THIS WEEK

EDITORIALS

CANCER Talk is so not cheap at the US National Cancer Institute **p.142**

WORLD VIEW The unlikely scientific wisdom of Chairman Mao **p.143**

NEWTS New proteins could help explain regeneration game **p.145**



The antibiotic alarm

There is a growing recognition that action must be taken to deal with the alarming rise in the incidence of bacteria resistant to today's antibiotics, and its implications for global health.

The threat posed by the spread of antibiotic-resistant strains of bacteria is unlikely to be news to *Nature*'s readers. Stories of the spread of 'superbugs' have become familiar in recent years. Yet what may come as more of a surprise is the relatively low level of recognition that this problem has received among policy-makers, especially when compared with the potentially catastrophic nature of the crisis that is unfolding. In essence, we are engaged in an arms race with pathogenic bacteria — and we are losing.

Academics, health-care professionals and campaigners have for decades warned that the misuse and over-prescription of antibiotics have led to a rise in resistant strains, a problem compounded by an alarming decline in the discovery and development of new classes of antibiotic. However, that message has grown stronger in recent years, and the previously disparate voices have begun to shout as one. Well-organized campaigns such as Antibiotic Action (a global initiative of the British Society for Antimicrobial Chemotherapy) have applied important pressure and raised the profile of the antibiotic-resistance threat with policy-makers around the world. At last, there are signs that the message is being heard.

This week saw the launch of a UK report into infections and the rise of antimicrobial resistance from Sally Davies, the UK chief medical officer. The report draws on the expertise of academics and health-care professionals to outline the burden of infectious disease in the United Kingdom and the increasing proportion of infections due to antibiotic-resistant strains. Davies makes 17 recommendations for policy and political action relating to antibiotic resistance, pathogen surveillance, prevention of infection and training for the health-care workforce. Chief among these recommendations is that antibiotic resistance should be added to the UK government's list of threats to national security, alongside pandemic influenza and terrorism, a recommendation alone that is sure to raise the profile of the issue. Importantly, Davies' report also recognizes the global nature of the challenge posed by antibiotic resistance and calls for international leadership and coordinated action. The report precedes the upcoming announcement of the UK cross-government anti-microbial-resistance strategy for 2013–18, which is supposed to make these things happen.

The call for better surveillance, better training for health-care professionals and more prudent use of antibiotics is also reflected in a second report, which was published on 5 March by the US Centers for Disease Control and Prevention and focused specifically on infections with carbapenem-resistant Enterobacteriaceae (CRE). Carbapenems are among the last options for treating many bacterial infections, and the increasing prevalence of CRE infections in the United States in the past decade is of concern, especially given that such infections are most often associated with patients undergoing serious medical procedures.

There have been many false dawns in the fight to raise the profile of antibiotic resistance and introduce meaningful policy changes. Despite expert advice to the contrary, antibiotics remain heavily over-prescribed by doctors around the world. In many countries they are available to buy over the counter without a prescription. In addition, in both the developed and the developing world, antibiotics remain in wide use as growth supplements in livestock, a practice that has drawn much criticism.

As the recent reports show, politicians and policy-makers are beginning to pay the antibiotic issue greater attention, and scientists should applaud them. Similarly, those individuals and groups that have campaigned to get us this far should be congratulated for their efforts. But nobody involved can afford to rest on their laurels. The next step must be to bring together all parties with a stake in the issue — researchers, health-care professionals and policy-makers, as well as representatives of the pharmaceutical and agricultural industries — to draw up a plan to tackle the crisis together. This will not be an easy process, and in addition to meaningful political will, it will probably require substantial funding. In the current global financial climate, this is likely to be a sticking point. But the potential financial and human cost of not tackling this crisis will almost certainly be much greater. It was not that long ago that loss of a sibling to a now-treatable disease was a common occurrence in Britain. We must not turn the clock back. ■

"Antibiotics remain heavily over-prescribed by doctors around the world."

Form and function

Although debate over scientific definitions is important, it risks obscuring the real issues.

Science is at the mercy of its language. It can be difficult for researchers to communicate what most excites them about the beauty, intricacy and complexity of the natural world. And when words fail, debates and arguments often arise.

One enduring debate has been resurrected by ENCODE, the Encyclopedia of DNA Elements — an ongoing multimillion-dollar project to catalogue the functional elements of the human genome. A headline-grabbing claim, first made in this publication last September, was that roughly 80% of human DNA had been ascribed some "biochemical function" thanks to the efforts of more than 440 scientists (The ENCODE Project Consortium *Nature* **489,** 57–74; 2012).

That percentage is remarkably high, in part because of a broad definition of 'function'. The ENCODE team used the term to include binding by a regulatory protein, or transcription into RNA — activities

© 2013 Macmillan Publishers Limited. All rights reserved

identified as widespread. But almost immediately, other scientists began to take this definition to task, calling it essentially meaningless.

Some background is useful. Genomes vary dramatically in size — sometimes irrespective of the complexity of the organism. Take, for example, the genome of the marbled lungfish (*Protopterus aethiopicus*), which clocks in at an excessive 133 billion base pairs. That of the pufferfish (*Takifugu rubripes*), by contrast, sports only 365 million.

For the ENCODE paper to suggest that humans have little genomic redundancy implies that the 3.2-billion-base-pair human genome hits a sweet spot in efficiency. Critics suggested, sometimes sharply, that this was both anthropocentric and ignorant of how evolution shapes the genome. Much of human DNA is non-functional, they insisted. It is a relic of history, garbled by mutation and essentially junk.

The most recent formal critique, published this week, suggests that similar analyses on organisms with very large and very small genomes would probably find the same density of functional elements (W. F. Doolittle *Proc. Natl Acad. Sci. USA* http://doi.org/kr3; 2013). This investigation has yet to be done.

The debate over ENCODE's definition of function retreads some old battles, dating back perhaps to geneticist Susumu Ohno's coinage of the term junk DNA in the 1970s. The phrase has had a polarizing effect on the life-sciences community ever since, despite several revisions of its meaning. Indeed, many news reports and press releases describing ENCODE's work claimed that by showing that most of the genome was 'functional', the project had killed the concept of junk DNA. This claim annoyed both those who thought it a premature obituary and those who considered it old news.

There is a valuable and genuine debate here. To define what, if anything, the billions of non-protein-coding base pairs in the human genome do, and how they affect cellular and system-level processes, remains an important, open and debatable question. Ironically, it is a question that the language of the current debate may detract from. As Ewan Birney, co-director of the ENCODE project, noted on his blog: "Hindsight is a cruel and wonderful thing, and probably we could have achieved the same thing without generating this unneeded, confusing discussion on what we meant and how we said it" (see go.nature.com/8xorge).

"To dismiss the debate as semantics minimizes the importance of words and definitions."

The ferocity of the criticism has no doubt been fuelled by dissatisfaction over ENCODE's top-down, big-science approach and the large share of research funds that it has attracted. Many biologists have called the 80% figure more a publicity stunt than a statement of scientific fact. Nevertheless, ENCODE leaders say, the data resources that they have provided have been immensely popular. So far, papers that use the data have outnumbered those that take aim at the definition of function.

The debate sounds like a matter of definitional differences. But to dismiss it as semantics minimizes the importance of words and definitions, and of how they are used to engage in research and to communicate findings. ENCODE continues to collect data and to characterize what the 3.2 billion base pairs might be doing in our genome and whether that activity is important. If a better word than 'function' is needed to describe those activities, so be it. Suggestions on a postcard please. ■

Cancer costs

Educating patients is key, but the US National Cancer Institute must keep spending in check.

Cancer experts from 15 countries last week published their reckoning of what it would take to combat the disease on a global scale. They highlighted the need for action and the scale of the problem — 12.7 million new diagnoses a year, more than half of them in developing nations — then ticked off a laundry list of developments needed to drive down that number, and to improve care for people with cancer.

It is an ambitious set of goals. From internationally agreed standards for preclinical animal models to harmonized clinical-trial protocols, the report lays out a framework that will require resources and coordination on an unprecedented scale (H. Varmus and H. S. Kumar *Sci. Transl. Med.* **5,** 175cm2; 2013).

The problem needs focus, and all involved must pull in the same direction. Yet last November, the same month that the cancer experts signed off on their recommendations, the US National Cancer Institute (NCI) resurrected a subcommittee of its advisory board to get to grips with a more local problem: the US$381.2 million that it had spent on its Office of Communications and Education (OCE) between 2006 and 2012.

The figure is high enough to make even bureaucratically hardened Washington DC insiders gasp. On 1 March, the respected cancer-research bulletin *The Cancer Letter* pointed out that the OCE's 2012 budget of nearly $45 million was almost double what the US Food and Drug Administration spent on communications, including drug and food safety announcements (P. Goldberg *Cancer Lett.* **39,** 9; 2013).

The news comes at an already troubling time for US cancer research. Funding agencies must strip 5% from their spending as part of budgetary sequestration, a failed political ploy to force lawmakers to trim the national budget. Where that will leave cancer research is still unclear: little information has trickled out to the public about how and where the cuts will be made. But the NCI's success rate for grant applications was already at 14%, an all-time low. At a press conference last September, director Harold Varmus predicted that the rate would fall even lower if sequestration were to take effect.

The OCE produces educational brochures for cancer patients and doctors, and runs Physician Data Query, a comprehensive database that includes clinical-trial summaries and definitions of medical terms. It also updates the NCI's websites, runs a hotline for patients and arranges the institute's exhibits at conferences.

In 2007, an external consultancy determined that the office had fallen prey to 'mission creep', losing its focus as a parade of directors shifted course repeatedly over the previous 10 years. Lack of focus, it seems, bred a behemoth. Whatever the cause of the largesse that allowed the OCE's bloated budget, it will only encourage lawmakers determined to slash access to public funds. And it is galling for those researchers who are scrambling to keep their labs alive. As *The Cancer Letter* has pointed out, the OCE's 2012 allotment would cover more than 100 coveted R01 research grants (P. Goldberg *Cancer Lett.* **38,** 45; 2012).

With the revival of its advisory board's subcommittee on communications, the NCI has an opportunity to end the mission creep, focus efforts and reduce its budget. This is not the first time that Varmus has had to trim the fat from NCI programmes. He successfully tackled the cancer Biomedical Informatics Grid, which came under fire in 2011 when NCI advisers questioned its $350-million price tag. And the institute has already taken steps to rein in the OCE's budget: its 2012 allotment, although large, reflected a 34% cut from 2006, when the office's spending topped $68 million. The subcommittee is next scheduled to present at an advisory board meeting in June.

There is no doubt that education of patients is crucial for cancer care and for clinical-trial recruitment. But the institute can surely continue to educate while tightening its belt, perhaps by consolidating the OCE's other administrative tasks. It must evaluate outside contracts and consider partnering with philanthropic groups to produce educational materials. In an era of ambitious goals and shrinking resources, that could free up much-needed money for research. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

© 2013 Macmillan Publishers Limited. All rights reserved



The unlikely wisdom of Chairman Mao

Self-criticism is a virtue seldom possessed by men, and never by the leaders of Western science, says **Colin Macilwain**.

Former Chinese leader Mao Zedong is remembered in the West as a monster. But he had at least one good idea — and it's one that those who speak for science should pick up on, and fast.

At the annual meeting of the American Association for the Advancement of Science (AAAS) in Boston, Massachusetts, last month, Rongping Mu of the Chinese Academy of Sciences ran through a searing litany of how China still lacks innovative capacity. As he did so, I thought of Chairman Mao, and his doctrine of self-criticism.

Instead of bragging about his country's growing scientific spending and prowess, Mu dwelt on its innovative weaknesses. He presented figures from the Organisation for Economic Co-operation and Development suggesting that China lags behind countries such as Russia and Turkey in its overall readiness to innovate.

What really caught my attention was that the guy from Japan stood up and did the same thing. Terutaka Kuwahara of the National Institute of Science and Technology Policy in Tokyo stressed deficiencies and weaknesses in the Japanese system. Ten years ago, you'd have been lucky to get this kind of self-criticism from a policy wonk speaking privately in Tokyo, never mind publicly in Boston.

Some time ago, the government of Singapore published a report that assessed the success and impact of various scientific initiatives in that country in pretty brutal terms. It found, for example, that the life sciences weren't creating jobs in the way that engineering was. A major revamp of the initiatives followed promptly.

One problem that Western science and technology faces is that this kind of analysis and action happens rarely in Europe, and never in the United States. It isn't that American culture is averse to self-criticism. In 1991, John Akers, then chairman of IBM, delivered a scathing talk, leaked to the press. I can still recall some of his words: "Too many people here hanging around the water-cooler…". Akers was later fired, but his harsh assessment helped prepare IBM to rebuild itself and to dodge the Internet iceberg that sank most of its old-time computing rivals.

And in 1995, a grizzled Motorola engineer called Bob Galvin was invited by the young, and possibly naive, Clinton administration to examine the performance of the US national laboratories run by the Department of Energy, NASA and the Department of Defense. His committee produced a scathing assessment. Galvin wanted Congress to pre-approve a commission that would assess, and close, some of these places. Since then, the report has gathered a lot of dust, and perhaps as much as US$100 billion has been poured down the drain to keep these labs open.

Nowadays, such critiques are notable for their absence. The argument, advanced at the AAAS meeting, that a heavily indebted and politically dysfunctional United States can shortly resume its twentieth-century growth pattern, if only it keeps investing in research and development (R&D), strains credulity.

In a session devoted to the connections between basic research funding and economic growth, no one produced a single paper or reference to support this argument. So I fished out some data of my own: in the year 2000, China spent about $25 billion on R&D and produced 3% of the global scientific literature. The United States spent around $300 billion on R&D and produced 27% of the papers. Thirteen years later, the Chinese economy has expanded by 200%; the American one by 20%. Clearly the two economies are not directly comparable, but curiosity-driven research on an industrial scale is a relatively recent invention, and I would suggest that it may be a sign — rather than the cause — of a successful economy.

I'm still waiting to see a PhD thesis — never mind a National Academies report — on the economic or health outcomes of the doubling of funding for the US National Institutes of Health from 1998 to 2003.

Over a pizza in Boston, I asked an old pal, who knows US science policy as well as any man alive, when he thought this troubling mist of complacency would lift. He offered the opinion that the pending money crunch and the economic challenge from China may soon prompt a change of tack.

As a Brit who spent 12 years in Washington DC, I'm not so sure. British scientists do well at getting papers published in high-profile journals, but this apparently impressive science base does not drive strong economic performance. Lately, Germany has enjoyed an export boom, whereas British exports have stagnated, despite a drastic 30% currency devaluation since 2008, because there is little productive capacity left in its economy. The leaders of British science, understandably, prefer to concentrate on its strengths rather than its weaknesses, and this is the trap into which the United States is also falling.

America needs to take a leaf from East Asia's book and take a cold, hard look at how it spends its money — including the vast federal R&D budget. As for Chairman Mao, he'd have cheered a report from the Xinhua news agency on 6 February. The story, "Frugality fuels science breakthrough", lauded a team of neutrino scientists for, among other things, finding cheap accommodation. I've never seen a story along these lines in Europe or the United States. And that, I'm afraid, suggests that decadence has set in. ■

AMERICA NEEDS TO TAKE A LEAF FROM EAST ASIA'S BOOK AND TAKE A COLD, HARD LOOK AT HOW IT SPENDS ITS MONEY.

➲ **NATURE.COM**
Discuss this article online at:
go.nature.com/vpl2ms

Colin Macilwain *writes about science policy from Edinburgh, UK.*
e-mail: cfmworldview@gmail.com

© 2013 Macmillan Publishers Limited. All rights reserved

RESEARCH HIGHLIGHTS

Selections from the scientific literature

PHYLOGENY

Many non-crop grasses bear salt

Although breeding salt-tolerance in crops is tricky, the trait has evolved many times in grasses — the family to which most crops belong.

Tom Bennett at the Australian National University in Canberra and his colleagues examined the family tree of 2,684 grasses and found that salt tolerance had independently evolved 76 times, indicating that its establishment does not require exceptional circumstances.

The authors suggest that the limited genetic diversity in modern crops could be a reason why breeding salt-tolerant varietals of cultivated plants has been troublesome.

Biol. Lett. 9, **20130029 (2013)**

CONDENSED MATTER

Atomic collapse on carbon sheets

The wonder material graphene has been used to confirm a long-standing prediction of quantum mechanics: that electrons in super-heavy atoms can spiral into the nucleus and away again, an effect known as atomic collapse.

Michael Crommie at the University of California, Berkeley, and his team assembled artificial super-heavy nuclei by depositing calcium ions on a graphene surface. Electrons behave as if they are massless in graphene's flat sheets of carbon atoms, and follow rules of relativistic quantum mechanics. This allowed the authors to detect the electronic signature of collapse for the artificial atoms using a scanning tunnelling microscope. The authors say that atomic collapse could one day be relevant for electronic devices.

Science **http://dx.doi.org/10.1126/science.1234320**



NIC MURRAY/ALAMY

BOTANY

Bacteria blossom in apple flowers

Surveys of microbe populations on plants have stuck mainly to leaves, but research now shows that flowers of apple trees (pictured) harbour a diverse and changing population of microbes.

Jo Handelsman at Yale University in New Haven, Connecticut, and her colleagues sequenced the DNA of microbes residing on the flowers of six apple trees at five points in their flowering cycle. In total, the authors identified 1,677 types of bacterium and archaeon.

From buds to petal-fall, six microbial groups colonized and dominated in succession. Although prevalent taxa produced the successional pattern, transient and rare taxa were behind tree-to-tree variation.

The team also found that the antibiotic streptomycin — which is used to control blight on commercial apple farms — lowers diversity, but does not affect the sequence of succession.

mBio 4, **e00602-12 (2013)**

CELL BIOLOGY

Membranes protrude to fuse

Whether it is an egg and sperm joining during fertilization or bone cells merging during development, cell fusion may start with a neighbourly poke.

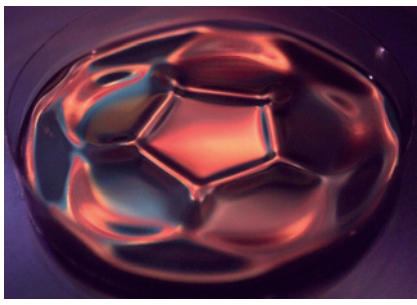
Elizabeth Chen and her team at Johns Hopkins University in Baltimore, Maryland, found a way to induce fusion in a cell line that does not normally fuse. When proteins involved in membrane fusion and cytoskeletal remodelling were expressed in these cells, the cells pushed finger-like appendages from their membranes by linking up subunits of actin — a protein required for cell movement and structure. This allows proteins at the tips of the 'fingers' to make direct contact with adjacent cells. The authors suggest that such actin-based appendages may be a general mechanism used by cells to initiate fusion.

Science **http://dx.doi.org/10.1126/science.1234781 (2013)**

WAVE DYNAMICS

Shaking oil into stars

By vibrating a shallow layer of oil, physicists have created standing wave patterns that



PHYS. REV. LETT.

© 2013 Macmillan Publishers Limited. All rights reserved

alternate between pentagons (**pictured**) and stars.

Jean Rajchenbach and his colleagues at the University of Nice Sophia Antipolis in France produced the patterns by shaking a 7-millimetre column of silicon oil up and down by up to 2 mm, between 7 and 11 times per second. The authors calculated that three separate surface waves passing through the oil interact to produce the shapes. The patterns created had triangular, pentagonal and hexagonal symmetries, depending on the frequency and amplitude of the vibration, but not the shape of the container holding the oil.

Phys. Rev. Lett. 110, **094502 (2013)**

NEUROSCIENCE

Human cells boost mouse brains

Mice performed better on learning tasks, such as navigating mazes, after receiving injections of human brain cells, whereas mice that received either no transplants or mouse brain cells showed no such improvement.

Steven Goldman and Maiken Nedergaard at the University of Rochester Medical Center in New York and their colleagues injected the brains of newborn mice with glial precursor cells, which go on to provide metabolic support to neurons and are thought to influence neuronal function.

When injected into the mice, the human precursor cells matured and integrated into mouse neural networks. Human glia strengthened neuronal signalling in mouse brains by secreting the cell-signalling protein TNF-α, which is thought to increase the number of receptors for the neurotransmitter glutamate. Such experiments, the authors suggest, can be used to explore the role of glia in human cognition and brain disease.

Cell Stem Cell 12, **342–353 (2013)**

HUMAN BEHAVIOUR

Victims punish but witnesses envy

Humans may be less inclined to punish bad behaviour than previous studies have suggested.

Michael McCullough of the University of Miami in Florida and his colleagues used computer-controlled games to look for evidence of altruistic punishment, in which an individual receives no clear benefit for inflicting punishment but does so anyway.

Although those who were treated unfairly in one game did punish transgressors in the next, the witnesses of the unfairness did not. In fact, witnesses were more likely to display envy of gains by others than outrage at victims' losses. The authors suggest that experiments to identify altruistic punishment are vulnerable to multiple artefacts, such as errors made by humans in predicting how they would react to hypothetical situations.

Proc. R. Soc. B 280, **20122723 (2013)**

METABOLISM

Ageing gene linked to diabetes

A gene called *SIRT1* is associated with age-related diseases and longevity in some model systems, but it seems that a mutation in this gene may also cause type 1 diabetes.

This form of diabetes is a result of the immune system destroying insulin-secreting cells in the pancreas. Marc Donath at the University Hospital Basel in Switzerland and his colleagues sequenced targeted regions of the genomes of a family in which four members have type 1 diabetes and another has ulcerative colitis, also an autoimmune disorder. The researchers found the *SIRT1* mutation only in family members with autoimmune diseases. Lab-grown cells that expressed the mutant gene boosted their production of the immune system components nitric oxide, chemokines and cytokines — all known to have a role in the development of diabetes.

Cell Metab. 17, **448–455 (2013)**

COMMUNITY CHOICE

The most viewed papers in science

MOLECULAR BIOLOGY

Regenerative proteins revealed

HIGHLY READ on genomebiology.com 20 Feb-13 Mar

Large-scale tissue profiling in the red-spotted newt (*Notophthalmus viridescens*, pictured) has revealed hundreds of new proteins, some of which could contribute to the newt's regenerative ability.

Newts can regrow severed limbs and some damaged organs, but the molecular processes driving this renewal have been hard to pin down — in part because the newt's enormous genome has not yet been sequenced.

Thomas Braun at the Max Planck Institute for Heart and Lung Research in Bad Nauheim, Germany, and his colleagues sequenced RNA transcripts from undamaged newt tissues, as well as tissue at different stages of regeneration, and hunted down the protein counterparts using mass spectrometry. Of the around 15,000 transcripts that the authors verified as protein coding, 826 coded proteins that are specific to newts. Some of those may represent new families of proteins.

Genome Biol. 14, **R16 (2013)**

For a longer story on this research, see go.nature.com/73sfqa



MYN/PIOTR NASKRECKI/NATUREPL.COM

PALAEONTOLOGY

Ancient camels in the Arctic

Whereas modern camels (*Camelus* spp.) live in hot, dry regions, their predecessors may have occupied polar forests.

Natalia Rybczynski at the Canadian Museum of Nature in Ottawa and her colleagues found fossilized fragments of a large leg bone (**pictured**) in the Canadian Arctic. Analysis of preserved protein showed that the bone belonged to an extinct giant camel. The remains dated to about 3.5 million years ago, a time when the region was densely forested and considerably warmer than today.

The fossil fragments, which are the northernmost evidence of camels, suggest that camel traits such as wide flat feet and even the iconic hump might have evolved as specializations for living not in the desert, but in the Arctic forest, say the authors.

Nature Commun. 4, **1550 (2013)**



NATURE.COM

For the latest research published by *Nature* visit:
www.nature.com/latestresearch

© 2013 Macmillan Publishers Limited. All rights reserved

SEVEN DAYS *The news in brief*

RESEARCH

Life in Lake Vostok

Russian scientists say that they have found a hitherto-unknown type of bacterium in Antarctica's largest subglacial lake. Samples retrieved last year from Lake Vostok contain an "unclassified" bacterium whose DNA is less than 86% similar to known bacteria, Sergey Bulat of the Petersburg Nuclear Physics Institute in Gatchina announced at a meeting last week in Moscow. More tests and cleaner samples are needed to establish the physiology and biochemistry of the microbe, says Bulat. See go.nature.com/ydcmw4 for more.

Carbon spike

Atmospheric carbon dioxide concentrations rose by 2.59 parts per million (p.p.m.) in 2012, marking the sharpest increase since 1998, according to data from the US National Oceanic and Atmospheric Administration. Atmospheric CO_2 concentrations reflect rising global emissions, driven by developing countries, as well as variations in carbon uptake by plants and the oceans. The global average, calculated from measurements at the Mauna Loa Observatory in Hawaii and other locations, exceeded 395 p.p.m. in January, representing an increase of more than 70% from pre-industrial levels.

Anti-HIV trial

A clinical trial of the preventive powers of the anti-HIV drug tenofovir may have failed because women were not actually taking the drug. In 2011, the VOICE study, in 5,029 HIV-negative women in South Africa, Zimbabwe and Uganda, suggested that neither a vaginal gel nor tablets containing the drug could prevent HIV infection (see *Nature* **480,** 10–11; 2011). Even though little unused product was returned, tenofovir was present in fewer than one-third of biological samples from participants assigned to receive it, study leaders revealed last week. Too few women took the drug to assess whether it did prevent HIV infection, although earlier studies have suggested it can work.



CIRO DE LUCA/INFOPHOTO/PHOTOSHOT

Rebuilding Naples' City of Science

The Italian government has pledged €20 million (US$26 million) to help rebuild the City of Science (*Città della Scienza*), an iconic 12,000-square-metre exhibition centre and science museum in Naples that was destroyed by fire on 4 March (pictured). Italy's research minister, Francesco Profumo, said that the government would work with local authorities on a plan to rebuild the museum in just 18 months. The museum's management is also seeking voluntary donations. Investigators had not announced an official cause for the fire as *Nature* went to press, but they were reported by the Italian media to suspect arson.

Elephant poaching

A surge in ivory poaching has driven the population of African forest elephants down to less than 10% of what could be supported by the available range. Researchers calculate that this species (*Loxodonta africana cyclotis*) lost 30% of its range and 62% of its population between 2002 and 2011 (F. Maisels *et al. PLoS ONE* **8,** e59469; 2013). The study was released as the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) met in Bangkok; many scientists have urged CITES to clamp down on poaching and habitat destruction and to tackle demand for ivory (see *Nature* **494,** 411–412; 2013).

Standard Higgs

The new particle discovered last year at CERN's Large Hadron Collider outside Geneva continues to behave just like the Higgs boson predicted by the standard model of particle physics, according to results presented last week at a conference in La Thuile, Italy. The latest data indicate that the boson decays into τ leptons as predicted, and also dampen earlier hints that the boson decays into pairs of photons more often than the standard model allows. No evidence yet points to theories beyond the standard model, such as supersymmetry (see *Nature* **491,** 505–506; 2012).

POLICY

Trading species

Polar bears (*Ursus maritimus*) will not be given increased protection under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), but it looks increasingly likely that several species of shark will. Delegates meeting in Bangkok for the sixteenth CITES conference, which governs the trade in many animals and plants, voted against banning the trading of polar-bear parts.

© 2013 Macmillan Publishers Limited. All rights reserved

GARY LEE/PHOTOSHOT

But a committee vote — which still has to be approved by the whole conference — did approve a need for export permits to protect oceanic whitetip (*Carcharhinus longimanus*), porbeagle (*Lamna nasus*) and other species of shark.

Quake appeal

Six scientists and a government official who were found guilty of manslaughter after the 2009 earthquake in the city of L'Aquila, Italy, have all filed appeals against the verdict in time for a 6 March deadline. Last October, a prosecutor sentenced each of the seven men to six years in prison, arguing that their reassurances led to the wrong messages being given to the public, adding to the death count in the earthquake (see *Nature* **490,** 446; 2012). Their sentences are suspended pending the outcome of the appeal. See go.nature.com/vny9un for more.

Antibiotic warning

The British government's leading medical adviser said on 11 March that bacteria that are resistant to antibiotics should be placed alongside pandemic influenza and terrorism on the list of national security threats. Sally Davies, England's chief medical officer (**pictured**), called for more international



research into new antibiotics and better surveillance of resistant microbes, in a report that charts the increase in drug-resistant bacterial infections. See go.nature.com/2u9ofh for more.

Cancer strategy

To fight the growing global burden of cancer, countries must boost research and develop national control plans for the condition, says a panel of experts from 15 countries. The report, published on 6 March (H. Varmus and H. S. Kumar *Sci. Transl. Med.* **5,** 175cm2; 2013), calls for a series of efforts including increased funding for cancer-prevention research, harmonization of international clinical-trial practices and global standards for biomedical research. The report's recommendations were the culmination of two meetings, the last in November 2012, of representatives from organizations that fund or conduct research. See go.nature.com/pxhoto for more.

BUSINESS

Drug spending

For the first time in more than two decades, the US public's spending on 'traditional' prescription drugs — therapies for common conditions such as raised cholesterol or high blood pressure — declined in 2012, according to a 5 March report by Express Scripts, a company in St Louis, Missouri, that manages pharmacy services. But the 1.5% drop in traditional prescription medications was more than offset by an 18.4% surge in spending for drugs that are used to treat complex diseases such as cancer and rheumatoid arthritis, and that often require special handling. Overall, US drug spending increased by 2.7%.

PEOPLE

Pasteur head

Christian Bréchot was on 8 March appointed the next director-general of the Pasteur Institute in Paris. Bréchot, currently vice-president for medical and scientific affairs at the Mérieux Institute in Lyons, was a former head of INSERM, the French national biomedical agency, but resigned in 2007 in relation to a legal case involving a cancer diagnostics firm that he and his wife had founded (see *Nature* **449,** 760–761; 2007). Bréchot succeeds Alice Dautry, who got the Pasteur Institute back on track after her predecessor had been evicted amid bitter staff disputes. Dautry had served the maximum two terms in office; Bréchot begins a four-year term on 1 October.

COMING UP

20 MARCH
The next winner of the US$3-million Fundamental Physics Prize is announced at CERN in Geneva, Switzerland.
go.nature.com/lr7hht

20–21 MARCH
In New York, researchers discuss the Sustainable Development Goals that will replace the United Nations' Millennium Development Goals in 2015.
go.nature.com/zwpgkh

18–22 MARCH
Results from the Mars Curiosity rover and the GRAIL mission to map the Moon's gravity are released at the Lunar and Planetary Science Conference in The Woodlands, Texas.
go.nature.com/eab6yw

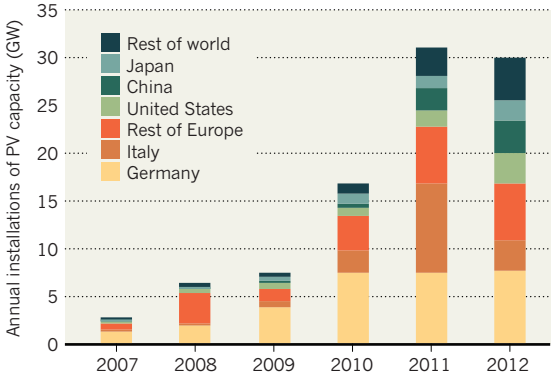
TREND WATCH

New connections of solar photovoltaic systems to the electricity grid fell in Europe for the first time last year, to 16.6 gigawatts (GW; see chart). Government withdrawals of solar subsidies helped to drive notable drops in Spain and Italy. But the non-European market expanded much faster, to more than 13 GW, meaning that the global total for 2012 reached 30 GW — not quite as much as in 2011, according to preliminary figures from the European Photovoltaic Industry Association in Brussels.

SOURCE: EPIA

SOLAR EXPANSION SLOWS

Cumulative global solar photovoltaic (PV) capacity reached more than 100 gigawatts (GW) last year — but annual installations fell.



CORRECTION

The item 'Go-ahead for iPS' (*Nature* **494,** 286–287; 2013) gave the wrong location for the review board that approved the iPS-cell study. It was the Institute for Biomedical Research and Innovation in Kobe.

NATURE.COM
For daily news updates see:
www.nature.com/news

© 2013 Macmillan Publishers Limited. All rights reserved

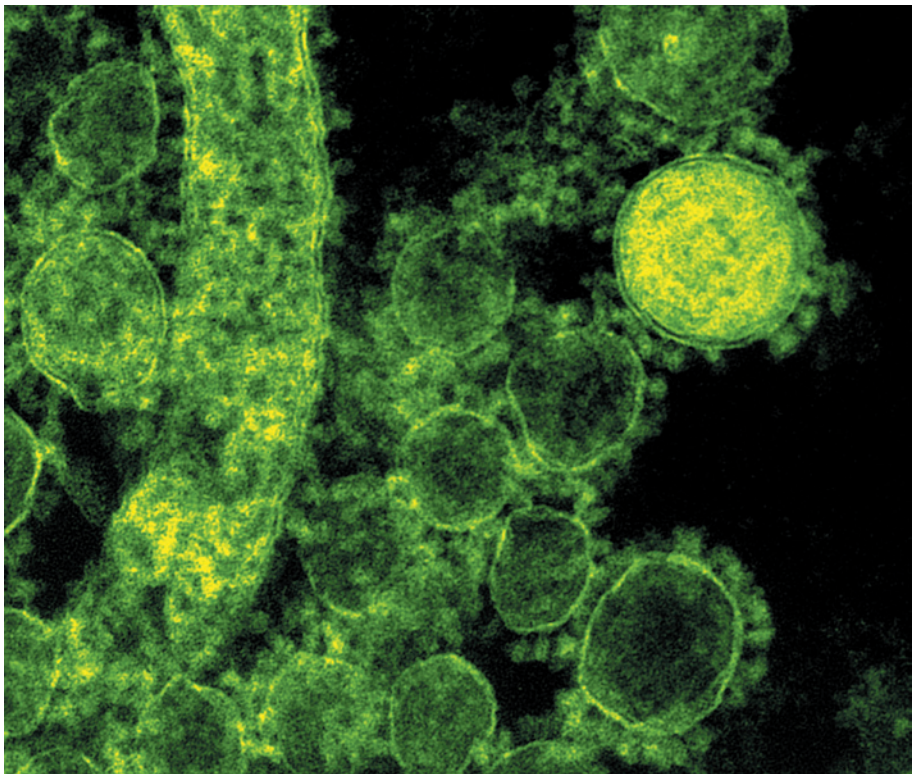
NEWS IN FOCUS

SYNTHETIC BIOLOGY One-stop shopping for gene controllers **p.150**

MATERIALS Silicene could be the next graphene — but not yet **p.152**

CONSERVATION Imperfect protection for Australia's marine riches **p.155**

EDUCATION Massive online courses bring big data to education research **p.160**



ELIZABETH R. FISCHER, ROCKY MOUNTAIN LABS/NIAID/NIH

Coronavirus hCoV-EMC is thought to have originated in bats.

VIROLOGY

Receptor for new coronavirus found

Virus might have many animal reservoirs.

BY DECLAN BUTLER

This week, researchers identified the molecule that has allowed a novel human coronavirus to infect at least 14 people, killing eight, since its detection last year. This key discovery, which pinpoints the receptor that the virus uses to infect cells in the human airways, opens up opportunities to study the virus's origin, the level of risk it poses and potential drugs and vaccines.

But it will take more than lab work to determine whether the virus is the next SARS — the coronavirus responsible for severe acute respiratory syndrome, which infected more than 8,000 people and killed more than 750 in the early 2000s — or just an exotic pathogen of little broad importance to public health. Only epidemiological data can show how efficiently the new coronavirus, hCoV-EMC, spreads from person to person and whether it is as deadly as it seems — such data are sorely lacking.

To jump to humans, animal viruses such as these novel coronaviruses, and avian and swine flu viruses, must evolve to be able to latch onto proteins on the surfaces of human cells. In a paper published this week in *Nature*[1], Stalin Raj at the Erasmus Medical Centre in Rotterdam, the Netherlands, and a largely European team report that spikes on the surface of hCoV-EMC bind to DPP4, a well-known receptor protein on human cells. When the binding site for the virus on DPP4 was blocked using antibodies, the virus could not infect cells; conversely, when DPP4 was expressed on the surface of normally non-susceptible cells, hCoV-EMC could now infect them.

The authors "unequivocally demonstrate that DPP4 is the receptor", says Ian Lipkin, a renowned virus hunter at Columbia University in New York. Although other receptors or co-receptors might be involved, the experiments suggest no need to look further, he says.

The work seems "solid and important", adds Michael Farzan, a virologist at the Scripps Research Institute's campus in Jupiter, Florida, who a decade ago led the team that discovered the receptor used by the SARS virus to enter human cells[2].

The receptor used by the SARS virus is found mainly on cells deep in the lungs. That helps to explain why that virus caused serious disease but was relatively difficult to catch or transmit by coughing or sneezing, notes Christian Drosten, director of the Institute of Virology at the University of Bonn Medical Centre in Germany and an author of the latest *Nature* paper. DPP4 seems to show a similar distribution, he says, which suggests that the new virus might behave in the same way.

But Stanley Perlman, a coronavirus researcher at the University of Iowa, Iowa City, and co-author of a related News & Views article also published in this week's *Nature*[3], cautions that knowledge of the distribution of DPP4 in the human airways is limited. The researchers did clearly show that the receptor is present on only some types of respiratory epithelial cells (about 20% of the total), which suggests that catching the disease might require inhaling high doses of virus, says Drosten. But another lab study, published last month in *mBio*[4], found that the virus infects human bronchial epithelial cells as easily as the coronaviruses that cause the common cold.

The nature of the DPP4 protein means that the virus could ▶

NATURE.COM
For more on the SARS virus see:
go.nature.com/wtadim

© 2013 Macmillan Publishers Limited. All rights reserved

remain a persistent threat. The receptor protein is present in very similar versions in many mammals, including bats, non-human primates, and various domestic animals. The virus might therefore easily jump between species, and humans might continue to be reinfected from a potentially wide range of animal reservoirs. This aspect of DPP4 is consistent with results published last year[5] showing that hCoV-EMC could infect bat, pig and human cells *in vitro*, an interspecies promiscuity not seen in the SARS virus or other coronaviruses.

From a public-health standpoint, it will be important to learn whether human cases of hCoV-EMC — so far largely centred on the Middle East — are caused by occasional jumps from animal reservoirs, or whether the virus has adapted and is now a distinct human virus spreading between people. The answer could dictate whether control measures should focus on human-to-human transmission or on transmission from livestock.

Tests in people who were in close contact with patients made ill by hCoV-EMC suggest that it does not, in fact, pass easily between humans. Volker Thiel, a coronavirus researcher at the Kantonal Hospital in St Gallen, Switzerland, and an author of the study[4] showing that the virus easily infects human cells, says that animal models are needed to better understand the factors affecting transmission.

Non-human primates are often used in coronavirus research, but researchers are also keen to study hCoV-EMC in more manageable species such as mice and ferrets. Bart Haagmans of the Erasmus centre, an author of the most recent *Nature* paper, says that this is difficult. Up to now he has been unable to infect ferrets, which he admits is "a bit surprising". It is unclear whether features of the virus or of the animal are the obstacle, he adds.

"What we need is classic gumshoe epidemiology."

Many of the human cases seen up to now cluster in families, which often indicates spread between humans but can also result from exposure to the same animal or environmental source. But cases in three members of the same family in the United Kingdom last month seemed to show unequivocal human spread. "What we need is classic gumshoe epidemiology," Lipkin says — to learn whether the virus is present in other animals, how prevalent it is, how people contracted it and whether it is shed into the environment, for example in faeces.

Genomic data from viruses isolated from infected people would also help, Drosten says. If the nucleic acid sequences of hCoV-EMC from different human cases are very similar, that might suggest an ancestral virus that has become an established human virus. But if the virus is only occasionally jumping to humans from an animal reservoir, one would expect to see far greater genetic diversity.

The biggest question is whether the novel coronavirus is truly the killer that the current data suggest — the mortality rate is more than 50% — or whether there are many undetected mild or asymptomatic cases. To answer this requires large-scale testing of the population, in particular of people living near outbreaks who have not fallen ill, to see if they have antibodies to the virus in their blood. That would indicate that they have been infected.

The assays needed have been developed in Drosten's and other labs, and Drosten insists that they are ready to use. Other researchers worry, however, that some tests might generate false positives by detecting other coronaviruses. Perlman says that it is urgent to start testing, especially in the Middle East. "This is the key issue." ■

1. Raj, V. S. *et al. Nature* **495,** 251–254 (2013).
2. Li, W. *et al. Nature* **426,** 450–454 (2003).
3. Gallagher, T. & Perlman, S. *Nature* **495,** 176–177 (2013).
4. Kindler, E. *et al. mBio* **4,** e00611-12 (2013).
5. Müller, M. A. *et al. mBio* **3,** e00515-12 (2012).

BIOTECHNOLOGY

DNA tool kit goes live online

Standard control sequences aim to make genetic engineering more predictable.

BY EWEN CALLAWAY

The latest shopping website is open for business, offering unusual wares: DNA tools to help biologists to engineer life.

The DNA sequences — which allow precise control of gene activity in the bacterium *Escherichia coli* — are the first output of BIOFAB, based in Emeryville, California, which calls itself "the world's first biological design-build facility". Launched in 2009 with a US$1.4-million grant from the US National Science Foundation, BIOFAB aims to advance synthetic biology by creating standard biological 'parts' in the form of DNA sequences that control gene expression. These standard sequences should allow biologists to engineer cells that can make medicines and perform other useful tasks simply by plugging in various sets of genes.

The sequences are meant to overcome a key barrier to synthetic biology: genes inserted into an organism do not behave predictably, even in such a well-understood workhorse as *E. coli*. "You would think after a generation of genetic engineering, expressing genes with precision in an organism as well utilized as *E. coli* would be pretty straightforward. It turns out it's not," says BIOFAB co-director Drew Endy, a synthetic biologist at Stanford University in California.



MARGOT HARTFORD

BIOFAB's directors Drew Endy (left) and Adam Arkin hope that their facility will help to industrialize synthetic biology.

For a cell to express a gene — that is, transcribe it into an RNA molecule and then translate that RNA into a protein — other sequences recognized by the cell's machinery must precede it. A promoter sequence is needed to make an RNA transcript, and a ribosome binding site (RBS) is crucial for protein translation.

Over the past three decades, scientists have amassed collections of these sequences and used them to express genes in which they are interested. Some sequences tend to be 'strong' and others 'weak', resulting in varying levels of RNA and protein being produced.

NATURE.COM
For more on synthetic biology, see:
go.nature.com/zbbytc

© 2013 Macmillan Publishers Limited. All rights reserved

But a team led by Endy and BIOFAB co-director Adam Arkin, of Lawrence Berkeley National Laboratory in Berkeley, California, has found that the activities of those sequences are far from predictable. In two papers published online this week in *Nature Methods*[1,2], the team reports inserting many different combinations of promoters and RBS sequences in front of genes encoding fluorescent proteins, and then measuring the level of protein that was made. "It was a bloody mess," says Arkin, with each promoter–RBS combination having varying effects depending on the gene.

He and Endy also cite an earlier finding that a scientist hoping to express a protein at a particular level has just a 50% chance of producing the required amount within a factor of two. Such hit-or-miss expression poses a major challenge to synthetic biologists who would like to create genetic circuits involving dozens of genes.

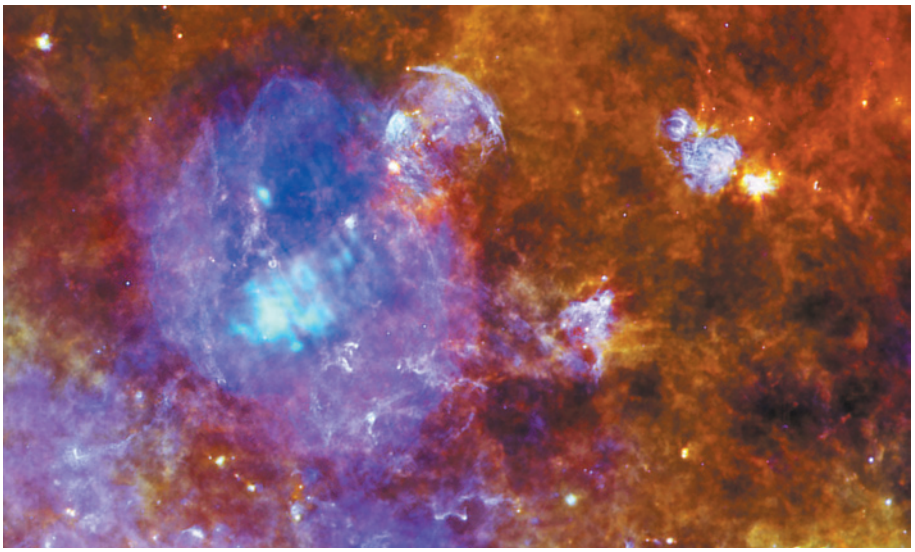
As a solution, the BIOFAB team designed promoter and RBS sequences for *E. coli* that do not interfere with downstream DNA, so that their effects are independent of the specific gene they are paired with. The sequences should provide scientists with a much tighter grip on gene expression, offering around a 93% chance of hitting a desired level of expression within a factor of two[2]. Researchers can obtain the sequences for free online (see http://www.biofab.org/data), and Arkin says that some of his colleagues are already finding them useful.

Endy and Arkin's team also devised a statistical method[1] to measure the variability in the performance of their promoter and RBS sequences, and indeed any genetic part to be used in synthetic-biology applications. The method should allow researchers to create a kind of specification sheet for each biological part, making it easier for scientists to develop and share their work.

Randy Rettberg, a synthetic biologist at the non-profit organization the iGEM Foundation in Cambridge, Massachusetts, who has worked with Endy on similar projects, says that more labs should follow BIOFAB's lead and industrialize the production of biological parts. And synthetic biologist Alistair Elfick of the University of Edinburgh, UK, says that the BIOFAB products should help synthetic biologists to design bigger and more complicated circuits.

"I think the community is very aware that we've got a long way to go before we can fulfil our dream of *in silico* design of genetic circuits we can just pop into a cell and run like an app," Elfick says. ■

1. Mutalik, V. K. *et al. Nature Meth.* http://dx.doi.org/10.1038/nmeth.2403 (2013).
2. Mutalik, V. K. *et al. Nature Meth.* http://dx.doi.org/10.1038/nmeth.2404 (2013).



HERSCHEL: Q. NGUYEN LUONG/F. MOTTE/HOBYS KEY PROGRAM CONSORTIUM/HERSCHEL SPIRE/PACS/ESA CONSORTIA. XMM-NEWTON: ESA/XMM-NEWTON

Herschel captured the shells of dust (orange) generated in supernovae (blue, from an X-ray image).

ASTRONOMY

Cold telescope faces hot death

Herschel space observatory nears its end after unravelling star formation and tracking dust from supernovae.

BY GEOFF BRUMFIEL

After more than three years of observations, astronomy's premier infrared space telescope is about to catch a fever and die. Later this month, Europe's Herschel space observatory, which has helped astronomers to revise theories about the birth and death of stars, will exhaust its stores of liquid-helium coolant, and its instruments will begin to heat up. At that point, "all the scientific instruments will shut down within hours", says Göran Pilbratt, the mission's project scientist at the European Space Research and Technology Centre in Noordwijk, the Netherlands.

Astronomers are hailing the legacy of the €1.1-billion (US$1.4-billion) mission, which has made some 22,000 hours of observations in the far infrared and submillimetre wavelengths, a part of the electromagnetic spectrum blocked by Earth's atmosphere. In an era when scientific spacecraft are increasingly specialized, the 3.5-metre Herschel telescope was a rare general-purpose observatory, used by more than 2,500 astronomers. "Anyone you ask who's been involved with Herschel has their own favourite results," says Matthew Griffin, an astronomer at Cardiff University, UK. "There's something for everybody."

Herschel orbits the L2 point, 1.5 million kilometres away in the cold shadow of Earth, where the combined gravity of the planet and the Sun create a 'gravitational well'. This shady perch, together with 2,300 litres of liquid helium, allowed Herschel to cool its instruments to a chilly 2.2 kelvin. At that temperature, the spacecraft could observe the low-temperature glow of gas and dust in stellar nurseries and in the shells of supernovae.

The cold Universe has held surprises. For example, astronomers thought that young stars form from long filaments of gas that collapse smoothly under their own gravity. Herschel painted a more complicated picture. When it looked at star-forming regions, it saw swirling, churning flows of gas driven by turbulent winds. Researchers now think that turbulence, rather than gravity, creates dense patches in the filaments that eventually collapse into stars, says Griffin. "That's a challenge for theoreticians."

The deaths of stars yielded other revelations. Astronomers had thought that most of the dust in the Galaxy forms in red giants, which puff it into space as they shrink in their waning years. Instead, Herschel detected massive amounts of dust ▶

➲ NATURE.COM
To see some images from Herschel, visit:
go.nature.com/aiylqs

© 2013 Macmillan Publishers Limited. All rights reserved

in the shock waves around supernovae, says Alexander Tielens, an astrophysicist at Leiden University in the Netherlands. "I think Herschel really nailed that supernovae make a lot of dust."

Outside the Milky Way, Herschel enabled observations of dusty galaxies from 10 billion years ago — when most of the Universe's stars were forming. The data show that stars tended to form evenly across these early galaxies, rather than being spurred by galactic mergers, says Gordon Stacey, an astronomer at Cornell University in Ithaca, New York. They also show that some giant black holes at the centre of galaxies, known as active galactic nuclei, hurl out jets of gas so powerful that they may prevent stars from forming in the vicinity. "It's pretty exciting to actually see these processes in action," says Phil Appleton, head of the NASA Herschel Science Center at the California Institute of Technology in Pasadena.

Herschel also allowed astronomers to look at a range of molecules in the Milky Way. Hydrogen fluoride worked as a tracer to reveal larger clouds of hydrogen gas, the building blocks of star formation. And water vapour turned up sometimes in unexpected places: stars made mainly of carbon and Jupiter's atmosphere, to name but two.

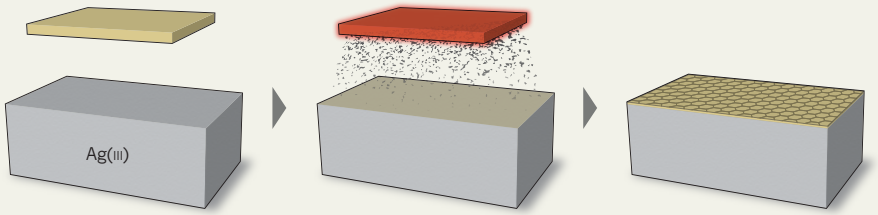
A new generation of instruments will follow up on Herschel's discoveries. From a perch in the Chilean Andes high enough to observe in the far-infrared, the Atacama Large Millimeter/submillimeter Array (ALMA) will point its dishes at distant galaxies first catalogued by Herschel (see page 156). The Stratospheric Observatory For Infrared Astronomy (SOFIA), a telescope carried by a high-flying 747 jet, will also be able to build on Herschel's observations. So will NASA's James Webb Space Telescope, scheduled for launch in 2018.

But "without Herschel there will certainly be a gap", says Stacey. Infrared astronomers want another space telescope that could make the same ultra-cold observations as Herschel, but with the sensitivity to reach farther into the Universe. To that end, the European Space Agency, which built Herschel, now hopes to collaborate with Japan to build the Space Infrared Telescope for Cosmology and Astrophysics (SPICA), a 3-metre-class telescope that would chill its mirror as well as its instruments. If the project wins funding, it could be launched sometime in the 2020s, says Pilbratt.

In May, after it shuts down, Herschel will be shunted to an orbit around the Sun to eliminate the risk of it falling back to Earth (an alternative plan to send it crashing into the Moon was abandoned owing to cost). But the observatory's public data archive will continue to lead to discoveries for years to come. "This is not the end of the mission," says Pilbratt. "This is the end of observing." ■

MAKING SILICENE

Atom-thick sheets of silicon — silicene — were first produced in 2010 but researchers have yet to grow the material on an insulating surface to test some of its predicted properties.



Step 1: Place silicon wafer above crystalline silver (Ag) in a vacuum.

Step 2: Heat the silicon until it begins to sublimate.

Step 3: Silicon atoms spontaneously organize on the silver surface as silicene.

MATERIALS SCIENCE

Sticky problem snares wonder material

Graphene-like form of silicon proves hard to handle.

BY GEOFF BRUMFIEL

In 2011, physicist Guy Le Lay stood before a half-filled room on the last day of the American Physical Society's March meeting in Dallas, Texas, and presented data on a new form of silicon. In his laboratory at Aix-Marseille University in France, Le Lay had grown sheets of honeycombed silicon with layers just one atom thick. He had only preliminary evidence that was unpublished at the time. "It was a risk, you know?" he says now of his decision to present the data.

At this year's meeting, on 18–22 March in Baltimore, Maryland, scientists will deliver about two dozen talks on silicene (see 'Speaking of silicene'), the material that Le Lay tentatively described two years ago.

The name recalls graphene, the current darling of the materials-science world — and the flurry of interest suggests that silicene could be the next one. But for that to happen, Le Lay and others will have to overcome silicene's unfortunate tendency to stick to practically everything it touches.

Structurally, silicene looks a lot like graphene, which is also a honeycombed sheet, but of carbon atoms rather than silicon. Silicene's two-dimensional structure should lead to strange quantum effects and allow electrons to streak across it at incredible speed — properties that, in graphene, have entranced physicists and builders of electronic devices since it was first characterized in 2004. In 2010, graphene's discoverers won a Nobel prize, and earlier this year, graphene research was selected by the European Commission as one of its billion-euro flagship projects (see *Nature* **493,** 585–586; 2013).

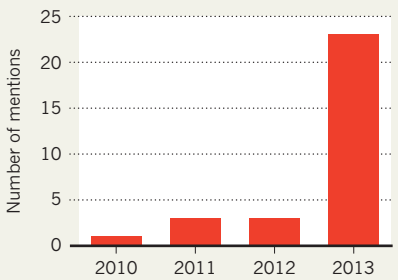
Silicene could even have some extra attractions. It is predicted to have characteristics similar to topological insulators — materials that conduct electrons only on their outer surfaces — another trendy area of research.

Above all, silicene is made of silicon, the same material that drives the modern electronics industry. Bringing it all together could lead to "a new era" in silicon electronics, says Kehui Wu, a physicist at the Chinese Academy of Sciences' Institute of Physics in Beijing.

There's just one problem: silicene is super sticky. "Graphene is a very stable material," says François Peeters, a condensed-matter theorist at the University of Antwerp in Belgium. But silicene reacts easily with the environment — oxidizing in the air and bonding chemically with other materials. And unlike graphene, which lies flat, silicene crinkles into bumps and ridges as a result of the way neighbouring silicon atoms bond with each other. That makes it more likely to stick to surfaces.

SPEAKING OF SILICENE

The number of times that silicene is mentioned in abstract titles for the American Physical Society's March meetings has shot up this year.



© 2013 Macmillan Publishers Limited. All rights reserved

Silicene's reactivity makes it much harder to produce than graphene. The Nobel prize-winning work on graphene began by peeling sheets from a block of graphite with a piece of sticky tape. Silicene, by contrast, can only be grown in an ultra-high vacuum on top of a material that matches its natural structure (see 'Making silicene').

Crystalline silver has proved to be the best fit because its atomic structure allows it to lock together with silicene's wavy ridges, and the silver's non-reactive surface means that it doesn't pull the silicene apart, Le Lay says. Reversing a technique he honed for depositing silver onto silicon, Le Lay grew the first samples of silicene on silver[1].

Only two other materials have been found to support silicene up to now. One, zirconium diboride, has the advantage of naturally sucking silicene onto its surface from a block of silicon positioned below[2]. The other, crystalline iridium, was reported as a possibility only in January this year[3].

Unfortunately, all three of these materials conduct electricity, says Yukiko Yamada-Takamura, a materials scientist at the Japan Advanced Institute of Science and Technology in Nomi. The bulky conductors mask silicene's delicate electrical properties, making it impossible to check whether the theoretical predictions of strange quantum effects are correct.

To see if silicene performs as expected, experimentalists will have to find a semiconducting or insulating surface on which to grow it. Better still would be to develop a technique to create free-standing sheets of silicene, Yamada-Takamura says. It's not entirely clear how that would be done, but given the increasingly competitive nature of the field, she says, "I will not tell you even if I had an idea."

As Peeters isn't racing to grow silicene himself, he's more willing to speculate. He thinks that sandwiching silicene between two sheets of another material, such as graphene, could stabilize it and prevent it from reacting with the outside world.

"I think if it will be used, it will be used in sandwich form, because that's the way in which you can stabilize it," he says. The outer sheets "can be any material; it really depends on what you want to do with it".

Despite its troubles, silicene's future looks bright. It has been included as part of Europe's massive graphene programme, and is catching on in the United States, Le Lay says.

The talks on it at the March meeting are likely to be more popular than they were two years ago, but Le Lay won't be there to find out. He's too busy giving seminars at departments everywhere from Hawaii to Austria to Japan. "It's crazy," he says. "But it's good!" ■

1. Vogt, P. *et al. Phys. Rev. Lett.* **108,** 155501 (2012).
2. Fleurence, A. *et al. Phys. Rev. Lett.* **108,** 245501 (2012).
3. Meng, L. *et al. Nano Lett.* **13,** 685–690 (2013).

MATERIALS SCIENCE

Exotic conductors from lab and nature

Mineral proves to be remarkably clean topological insulator.

BY ZEEYA MERALI

They say that it's what's on the inside that counts. But that is not true for topological insulators — exotic materials that conduct electricity only along their surfaces. A team of physicists has now demonstrated this property in a naturally occurring mineral[1], and another group has synthesized the first two-dimensional topological insulator that conducts at room temperature[2].

Having a broader range of such materials could boost researchers' efforts to build spintronic devices — in which currents are driven by an intrinsic property of electrons called spin, rather than by voltages. The materials could also help the design of quantum computers that would use spin to encode information.

Predicted to exist in 2005 (ref. 3), topological insulators that work at low temperatures were first synthesized from heavy elements in 2008 (ref. 4). Their odd conducting abilities arise because each electron's spin becomes coupled to its motion. This relationship compels each electron to circle around a specific spot, preventing them from moving through the bulk material, which means that they cannot conduct electricity. But at the material's edge, the electrons do not have enough space for this circling motion; instead, they are forced to hop along the surface in semicircular jumps, enabling conduction.

The thin conducting layer of a topological insulator makes it relatively easy for physicists to manipulate the spin current. "Topological insulators raise the possibility of building spintronic devices that use electron spin, rather than charge," says Pascal Gehring, a solid-state physicist at the Max Planck Institute for Solid State Research in Stuttgart, Germany, and a co-author of the mineral study. Spins can be rotated quickly without expending much energy, so spintronic devices should be more efficient than their electronic counterparts, in which energy is required to change charges, he adds.

Physicists attempting to construct quantum computers that would outperform the best current machines are also interested in encoding information in electron spins. In theory, it is difficult to corrupt spin values in a topological insulator. That is because, to flip the spin value accidentally, you would have to knock the system hard enough to cause the electron to make a complete U-turn.

In search of materials that display these properties, Gehring and his colleagues examined a natural sample of kawazulite, which contains bismuth, tellurium, selenium and sulphur, found at a former gold mine in the Czech Republic. Lab-made samples of kawazulite have already been shown to be topological insulators, but no one had checked for the property in natural samples.

"It may turn out to be cheaper to use a natural supply."

The team cleaved off single crystalline sheets 0.7 millimetres wide and applied the standard test for a topological insulator: photoelectron spectroscopy. This involves measuring the properties of electrons dislodged when ultraviolet light is fired at a material's surface. Their results[1] confirm that the electrons' energy and momentum distribution matches predictions for a topological insulator.

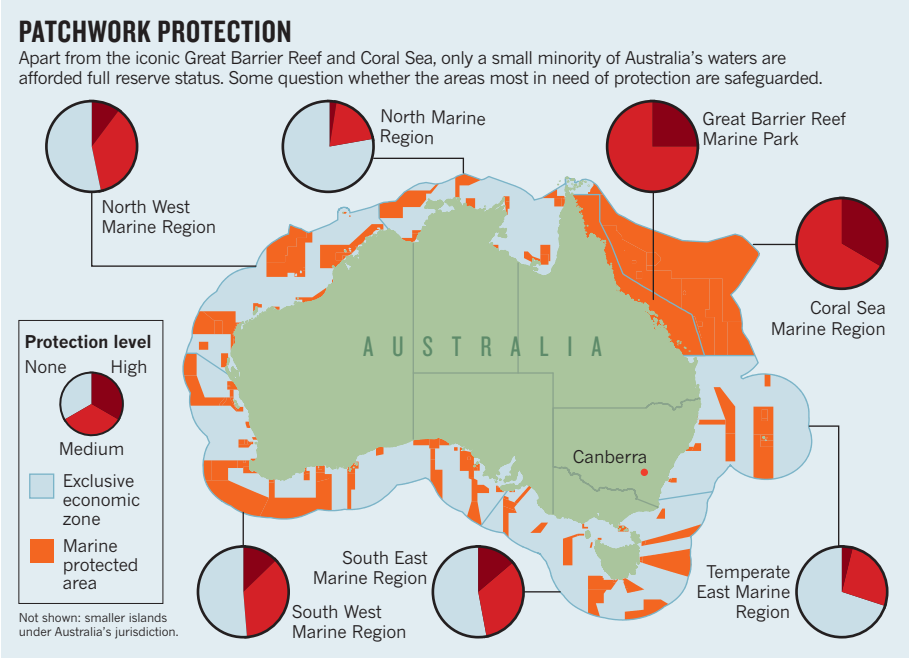
Feng Liu, a materials scientist at the University of Utah in Salt Lake City, notes that the team's natural sample contains fewer structural defects than its lab-made counterparts, reducing unwanted conduction in the bulk. "It may turn out to be cheaper to use a natural supply of topological insulators," says Liu.

Even in the lab, topological insulators require less exotic conditions than had been thought. Jeroen van den Brink, a physicist at the Leibniz Institute for Solid State and Materials Research in Dresden, Germany, and his colleagues stacked bismuth-containing sheets with a honeycomb structure like that of graphene. The result is a bulk material that acts as topological insulator at room temperature[2].

The next step should be to find organic materials that act as topological insulators, says Liu. His team recently proposed a design for such a compound[5], and says that another group has synthesized a candidate structure. "Ultimately, these will be these cheapest and most versatile materials to work with." ■

1. Gehring, P. *et al. Nano Lett.* http://dx.doi.org/10.1021/nl304583m (2013).
2. Rasche, B. *et al. Nature Mater.* http://dx.doi.org/10.1038/nmat3570 (2013).
3. Kane, C. L. & Mele, E. J. *Phys. Rev. Lett.* **95,** 146802 (2005).
4. Hsieh, D. *et al. Nature* **452,** 970–974 (2008).
5. Wang, Z. F., Liu, Z. & Liu, F. *Nature Commun.* http://dx.doi.org/10.1038/ncomms2451 (2013).

© 2013 Macmillan Publishers Limited. All rights reserved



CONSERVATION

Australia's plans for sea havens 'flawed'

Questions raised over protection levels for marine regions.

BY DANIEL CRESSEY

When Australia unveiled a groundbreaking network of marine reserves to gird the continent last year, praise rained down from conservationists. The plan, the most ambitious yet in the world, was seen as a secure, robust way to safeguard biodiversity, adding new protections for unique habitats such as the less-studied northern waters to existing ones for the iconic Great Barrier Reef.

But as scientists probe the details, they are finding worrying shortcomings, raising doubts about the rush to create similar marine protected areas elsewhere in the world.

Australia's government will spend tens of millions of dollars on the programme. It will protect 2.3 million square kilometres of sea (see 'Patchwork protection') and try to balance leisure uses, commercial interests such as fishing and drilling, and conservation. The plans, put out for consultation last June, became official in November. Parliament is expected to approve the last step in the process shortly.

The government decided which activities should be allowed in which areas after examining scientific information on the habitat types and soliciting comments from local residents, industry representatives and conservationists. Its official guidelines said that the result should be "comprehensive, adequate and representative", sustaining diversity while allowing tourism, shipping and resource extraction.

But an assessment of the scheme published late last month (L. M. Barr and H. P. Possingham *Mar. Policy* **42,** 39–48; 2013) revealed flaws in the proposed network, which includes reserves in which all fishing is banned, as well as areas in which some fishing and other activities are permitted.

The authors of the assessment — Lissa Barr and Hugh Possingham of the University of Queensland's Centre of Excellence in Environmental Decisions in Brisbane — looked at how representative the proposed areas are. Protecting the full range of ecosystems off the coasts of Australia is a "fundamental requirement" for a good reserve network, they say, or else habitats with distinct biology may be lost.

What they found was troubling. Except for the high-profile Great Barrier Reef and Coral Sea regions, the protected areas are "far from representative", they wrote. Of the 85 "bioregions" that the Australian government identified within its waters, 30 have the highest level of protection for less than 1% of their area. In four cases, less than 1% is covered even when areas of higher and lower protection are added together.

A key problem is that the highly protected areas tend to be in deep water, whereas the most vulnerable ecosystems are usually closer to shore, on the continental shelf, says Possingham. He notes that many of the shallower areas in the north and northwest, which are underrepresented in the plans, are suspected to have fossil-fuel potential. "One could wildly speculate that the Australian government is reluctant to forgo opportunities in that area," he says.

The Australian Department of Sustainability, Environment, Water, Population and Communities, which oversaw the creation of the network, said that it would not be able to comment before *Nature* went to press.

Bob Pressey, who works on conservation planning at James Cook University in Townsville, is also concerned. He says that, just like land-based reserves, the regions chosen by nations to be marine reserves, especially around Australia, tend to be the least commercially useful and so the least controversial to conserve, even though they may also be the least threatened. "This is a problem that we should be worried about," he says.

Countries are keen to learn how best to set up marine protected areas so that they can comply with the Convention on Biodiversity, an international agreement specifying that countries should be conserving 10% of their marine regions by 2020 (see *Nature* **480,** 166–167; 2011). "We're watching it [Australia] very closely," says Les Kaufman, a biologist at the Boston University Marine Program in Massachusetts. He says that similar exercises are under way in California and New England, and that lessons so far from these and other regions suggest that at least 10%, and preferably 20%, of each habitat must have the highest levels of protection to truly safeguard all the ecosystems.

Despite their concerns, experts say that the trend to conserve ecosystems, rather than just individual species, is encouraging. "What Australia has attempted here is a world first," says Imogen Zethoven, who leads work on Australia's marine systems for the Pew Charitable Trusts environmental non-governmental organization in Sydney. "The proposed reserves are a very big step forward for marine conservation."

Possingham, too, praises Australia's effort, saying that "it's a really good first step", but warns that "nobody should imagine the job done". ■

CORRECTIONS

The News story 'Cholesterol limits lose their lustre' (*Nature* **494,** 410–411; 2013) gave the wrong location for Decision Resources; it is in Burlington, Massachusetts. And the iPS-cell study mentioned in the News story 'Stem cells cruise to the clinic' (*Nature* **494,** 413; 2013) was not approved by the review board at Japan's Center for Developmental Biology but by the central review board at the centre's parent organization, RIKEN in Wako.

© 2013 Macmillan Publishers Limited. All rights reserved

THE PATCHWORK ARRAY

After years of delays and cost overruns, an international collaboration is finally inaugurating the world's highest-altitude radio telescope.

BY ERIC HAND



Eyes on the sky at the Atacama Large Millimeter/submillimeter Array.

STÉPHANE GUISARD/ESO

The car toils upwards along the sinuous road, its engine tuned for the thin air. The clumps of cactus and grass along the road soon give way to bone-dry lifelessness. By the time the car reaches 4,000 metres above sea level, Pierre Cox has a bit of a headache. By the time it reaches the 5,000-metre-high Chajnantor plateau — one of the highest, driest places on Earth, and one of the best for astronomy — the altitude is affecting his bladder. Cox, the incoming director of the Atacama Large Millimeter/submillimeter Array (ALMA) in Chile, is about to glimpse the giant telescope dishes he will soon be responsible for. But first he must find a toilet.

Cox slides out of the car and staggers into ALMA's glass and steel operations centre. The current director, Thijs de Graauw, a trim 71-year-old Dutchman, follows Cox inside and sits down. For him, journeys like this occur weekly — if not daily — but he knows that they are no joke. First-timers get a mandatory medical screening before being allowed up to the plateau, and regular shift workers pad around the building with tubes in their noses and oxygen tanks on their backs. "Everyone okay?" De Graauw asks the group of astronomers who have accompanied Cox to ALMA on this December day. "No victims yet?"

Cox re-emerges from the toilet, puts on wraparound sunglasses and, slightly dizzy, heads outside with the group. Scattered across the surrounding plain of brown volcanic soil are dozens of huge white radio antennas, looking as out of place as the stone statues on Easter Island. High on this cold and lonely plateau, they are gathering photons from the cold and lonely parts of the Universe — the dimly glowing clouds of dust and gas where stars are born. Their signals are then combined into images that have a resolution better than that of the Hubble Space Telescope.

The stillness of the tableau breaks as the dishes begin to tilt and swivel in unison. "My goodness," says Cox, hushed by the sight of so much metal moving so quickly and quietly.

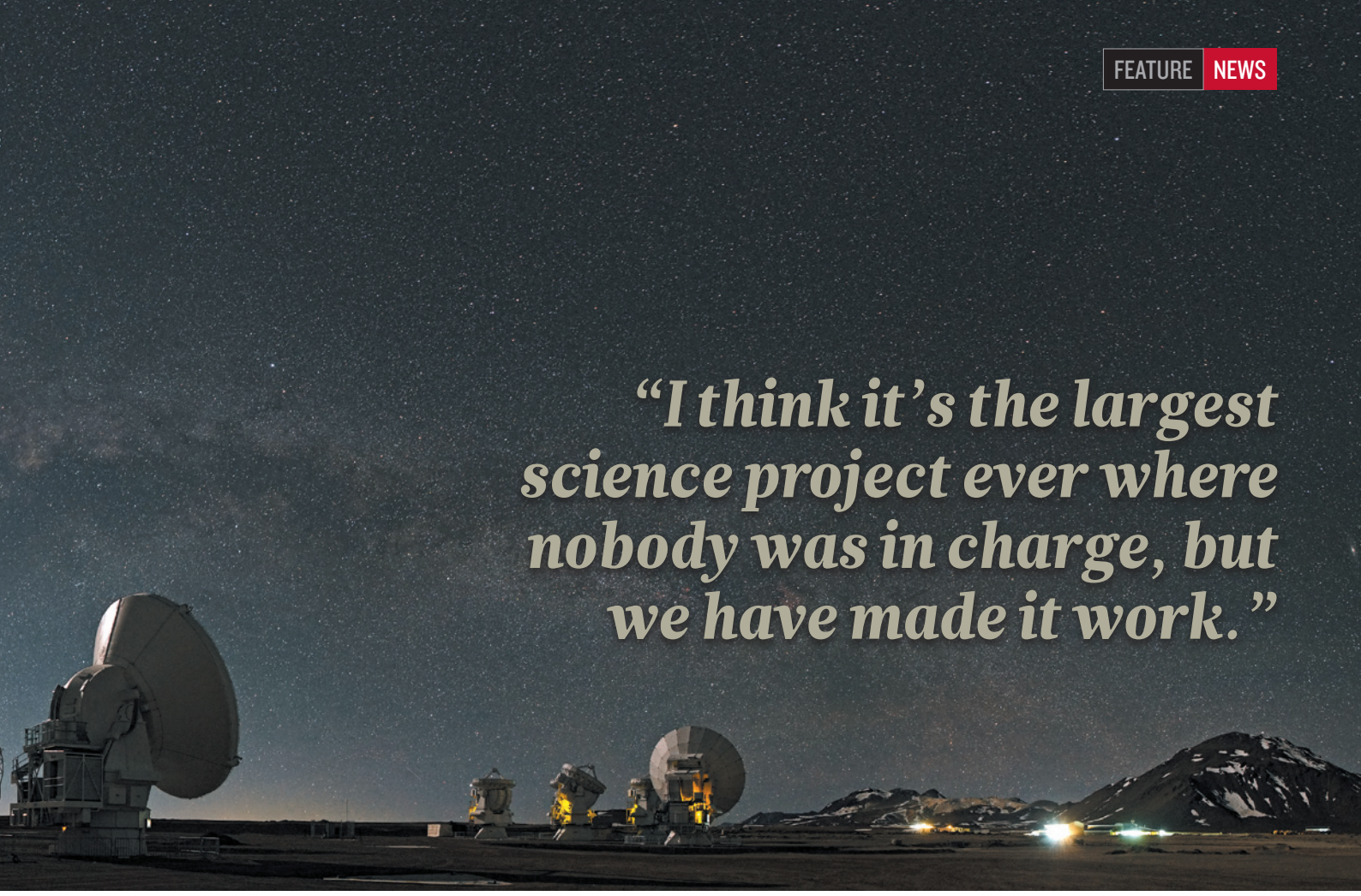
But the choreography is not quite uniform. Clustered tightly in the middle of the array are 12 dishes, each 7 metres across, and four 12-metre dishes, from Japan. Spaced farther out are 25 dishes, each 12 metres across and fitted together like pie slices, from the United States. And scattered among those are the first of 25 dishes from Europe, each 12 metres across — top-of-the-line carbon-fibre devices pivoting on silky-smooth gearing.

The last of those European antennas will not be installed until the end of 2013, when ALMA will finally reach its full complement of 66 dishes. Rather than wait until then, however, the project held a formal inauguration ceremony on 13 March to celebrate the collaboration that made it all possible. A total of 19 countries have contributed to ALMA, through three primary partners: the European Southern Observatory (ESO); the National Astronomical Observatory of Japan; and the US National Radio Astronomy Observatory (NRAO) in Charlottesville, Virginia, funded by the US National Science Foundation (NSF).

Less celebrated have been the difficulties of keeping this unwieldy confederation on track — with power shared among three independent organizations that have different cultures and norms. Nor is anyone likely to cheer about how the lack of unity caused the US$1.4-billion project to come in several years late, well over cost and downsized from its original ambitions. Its successive directors have had to be diplomats and negotiators as much as scientists.

But ALMA is not unique in that respect. International mega-projects are becoming increasingly common in astronomy. Witness the Square Kilometre Array, a proposal to build 3,000 radio dishes with a total collecting area approaching 1 square kilometre in Australia and South Africa (see *Nature* **484,** 154; 2012). As costs for such ambitious projects cross the billion-dollar threshold, nations are finding that they cannot go it

© 2013 Macmillan Publishers Limited. All rights reserved



alone — a situation for which ALMA might serve as a valuable object lesson. “I think it’s the largest science project ever where nobody was in charge,” says Ethan Schreier, president of Associated Universities Incorporated (AUI), a radio-astronomy research-management company based in Washington DC, which operates the NRAO. “But we have made it work.”

FAMILY TIES

Each of the three primary partners came to Chile in its own way, with pilot projects dating from the 1980s. Of particular interest for Europe was the infrared glow from the dust that shrouds many of the Universe’s first galaxies. This glow can be used to estimate the size, brightness and number of stars hidden within — key questions for astronomers trying to piece together the history of galaxy formation. Shifted to longer wavelengths by the expansion of the Universe, this glow reaches telescopes on Earth as millimetre-wave radiation — and can be detected day or night, as long as there isn’t much atmospheric water vapour in the way.

To get at the earliest (and thus most distant and faint) of these dusty galaxies, European astronomers needed a large collecting area. They proposed an array of 16-metre dishes on the salt flats of the Atacama Desert, more than 1 kilometre lower than the Chajnantor plateau. US astronomers were more interested in star formation within our Galaxy, and wanted the better image quality that would come with an array of 8-metre dishes placed more closely together. They also wanted to push into the shorter wavelengths of the unexplored submillimetre band, where they could study chemical-emission lines from molecules in interstellar gas clouds. They pushed hard for Chajnantor, which was high and dry enough for submillimetre observations, and flat enough for the dishes to be moved into various configurations.

Pooling resources was an obvious move for the two projects, and in 1997, ESO director Riccardo Giacconi and NRAO director Paul Vanden Bout signed a joint resolution to pursue a compromise — a facility of 64 dishes, each 12 metres across. “Riccardo and I signed this document with no authority whatsoever,” says Vanden Bout. The official backing from the ESO and the NSF wouldn’t come for another six years.

Japan joined the partnership in 2004 and committed to building 16 dishes in the centre of the array. The more widely spaced US and European dishes would provide high-resolution detail in a narrow field of view. But the compact array would give a more complete view of large objects such as the Galactic Centre or the sprawling, dusty clouds where the Milky Way forms its new stars.

The patchworked nature of ALMA’s creation is reflected in its organizational structure. Partner agencies have been loath to relinquish control over budgets (or anything else), so the coordinating body that manages array operations — the Joint ALMA Observatory (JAO) — has no formal authority. For example, when Chile created a science preserve on the Chajnantor plateau (in return for 10% of ALMA’s observing time), officials signed the lease with AUI and the ESO, not the JAO.

CULTURAL SENSITIVITY

ALMA directors quickly learn that management works best through persuasion, not proclamation. “You have to seduce,” de Graauw says. Cultural sensitivity is also required. On conference calls, de Graauw says, his Japanese colleagues would say nothing until he solicited them directly for comments. Alison Peck, deputy project scientist for ALMA, learned a similar lesson about setting deadlines. “In Japan, it’s really not okay to miss a deadline,” she says. “In the United States, you can usually make reasonable excuses and ask for an extension. In Europe they worry about it even less.”

ALMA’s motley nature is apparent even in the 12-metre telescope dishes, the array’s biggest single cost. From the beginning, the technical requirements were “truly daunting”, says Tony Beasley, a former ALMA project manager and current head of the NRAO. Each dish needed a motor that could accurately point at celestial targets to within

© 2013 Macmillan Publishers Limited. All rights reserved

0.6 arcseconds (about the same apparent size as a bacterium at arm's length); a reflecting surface with an accuracy of 25 micrometres (about one-quarter of the width of a human hair); and structural materials that could maintain that precision in the face of Chajnantor's wicked winds and subzero temperatures.

The cheapest way to meet those requirements would have been for the ESO and the NRAO to share a single design and a single contractor. But the NRAO went with a small US firm — Vertex, which was later bought by General Dynamics of Falls Church, Virginia — and the ESO held out for a European consortium led by Thales Alenia Space, based in Paris. The delays associated with going to separate contracts came just as prices for commodities such as steel were rocketing because of demand in China, leading to a dramatic escalation in ALMA's cost. As a result, in 2005, the project was 'descoped' — the NRAO and the ESO would each contribute only 25 antennas rather than 32, resulting in a loss in array sensitivity (see *Nature* **439,** 526–528; 2006). Even with the descope, US and European contributions to ALMA would grow from $650 million to $1 billion.

Japan, meanwhile, had contracted its dishes to Mitsubishi Electric, based in Tokyo. The three companies maintained separate assembly sites at the ALMA operations support facility (OSF), a cluster of buildings where most staff members live and work. (The OSF was built at 2,900 metres, in part because it costs less to hire Chilean workers for altitudes lower than 3,000 metres.)

It is too early to tell whether one design will outperform the others. The ESO's carbon-fibre dishes change pointing position with fewer errors, but it is uncertain how well the advanced internal gearing will hold up to weather over time. So far, all the antennas are performing to specifications. But having three different designs will saddle ALMA with extra operations costs far into the future, says Neal Evans, an astronomer at the University of Texas at Austin and chair of the ALMA board. "You'll need different spare parts, and you'll need people that know how to maintain each of the designs," he says.

AMBITIOUS TARGETS

Despite all the headaches, antennas are steadily accumulating on the plateau. In 2007, the JAO team raised glasses of water to celebrate the first linking of two dishes using the correlator — a computer that connects dish signals to create a composite view of the sky. (Why no champagne? The altitude impairs judgement, even at 2,900 metres, so ALMA has a strict no-alcohol policy; workers are subject to random breathalyser tests on the buses connecting local towns to the OSF.)

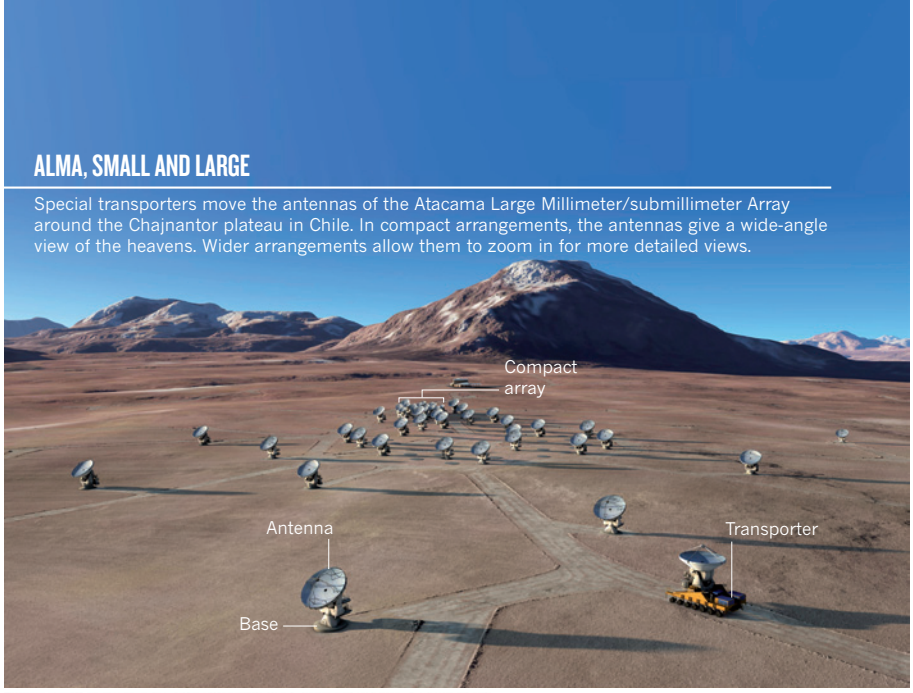
In September 2011, with 16 dishes in place, ALMA began its inaugural observing period with the 100 or so projects that had risen to the top of its 'cycle 0' proposal competition. Most of the observation targets were relatively nearby objects in our Galaxy. Results ranged from the detection of sugar-related molecules in a nearby star system to an exceptionally sharp image of the gas clumps that will collapse into giant stars (see *Nature* **492,** 319–320; 2012).

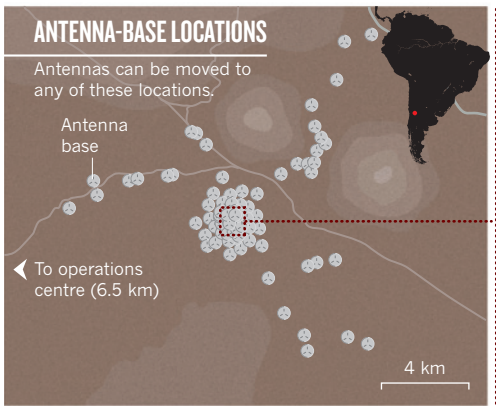
But the targets will soon become more ambitious. The mathematics of radio arrays implies an inverse relationship between antenna spacing and image resolution: the longer an array's 'baselines' (the distances between pairs of antennas), the smaller its field of view and the higher its resolution. The number of baselines, which determines how 'filled in' an ALMA image is, has grown simply through the addition of antennas. But the observatory can also change baselines by moving the antennas around the plateau — with the help of two German-built transporters nicknamed Otto and Lore (see 'ALMA, small and large').

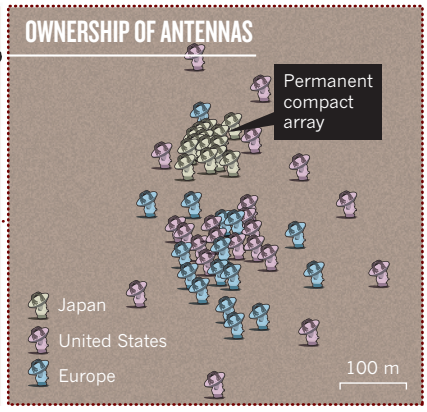
In January, ALMA began its cycle 1 observations with 32 of the 12-metre dishes working at baselines of up to 1 kilometre, combined for the first time with some of the smaller Japanese dishes in the centre of the array. By the time cycle 2 begins, in early 2014, ALMA astronomers hope to have 40 dishes working at baselines of up to 2 kilometres. The resulting high resolution will help astronomers to understand star formation in distant galaxies seen very early in their lives, when the Universe was young and its chemical composition was different. "ALMA could very well open up a whole new field of star formation," says Linda Tacconi, an astronomer at the Max Planck Institute for Extraterrestrial Physics in Garching, Germany.

ALMA will also be able to pinpoint how far away, and therefore how old, an object is. Usually, that measurement is a two-step process. Researchers first need time at a radio-astronomy facility to locate the object — a distant galaxy, say — then must spend many hours on an optical telescope to split the faint light up into its spectral components and identify emission or absorption lines caused by the presence of various elements and molecules. Measuring how far those wavelengths have been stretched by cosmic expansion allows observers to estimate how far away the object is.

ALMA can do all of the above within minutes. Already, ALMA observations have shown that strangely bright early galaxies were in fact multiple smaller galaxies that had been lumped together by an earlier optical survey (A. Karim *et al.* Preprint at http://arxiv.org/abs/1210.0249; 2013). The discovery was a relief to theorists, who had been unable to work out how such bright, huge galaxies could have formed so early in the Universe.







ALMA (ESO/NAOJ/NRAO)/L. CALÇADA (ESO)

NATURE.COM
To read more about science at ALMA, visit:
go.nature.com/cv5btp

© 2013 Macmillan Publishers Limited. All rights reserved



The Antennae galaxies as observed by ALMA (red and yellow) and the Hubble Space Telescope.

ALMA (ESO/NAOJ/NRAO); NASA/ESA HUBBLE SPACE TELESCOPE

Once ALMA reaches baselines of 10 or more kilometres, astronomers will be able to turn their attention to stars forming in our Galaxy. The observatory has already detected gas flows in the disk surrounding a newborn star, crossing a gap that indicates the presence of a giant planet (S. Casassus *et al. Nature* **493,** 191–194; 2013). Eventually, for some of the star systems closest to Earth, ALMA astronomers could have a shot at seeing the whirlpools of gas in which planets themselves are coalescing.

THE DEEP UNKNOWN

But most of these projects will have been preordained — interesting stars, clouds or galaxies already seen in different parts of the spectrum by other telescopes. Many astronomers think that ALMA needs to forge a new path. They are calling for a 'deep-field survey' — a time exposure of a patch of sky for hundreds of hours, long enough to image extremely faint objects in that field and, possibly, to glimpse the formation of the Universe's first galaxies. "I think it's something that has to be done," says Leonardo Testi, ALMA project scientist for the ESO. "If you only follow up on something else, then you are only looking after things you already know."

The question is whether the JAO is strong enough to marshal ALMA's partners to do a deep-field survey. The Hubble Space Telescope has done several such surveys over the past two decades — but only because Hubble directors have allocated large chunks of discretionary time to the projects, thereby circumventing the fierce competition for observation slots. ALMA directors have very little discretionary time — almost every data-taking moment has been allocated. To do a deep-field survey, the partners would have to donate the time — a tough sell for a facility now receiving around six proposals for each available slot.

The problem highlights a complaint common among JAO staff: none of the partners can call the shots. Europe and the United States have equal shares, both larger than Japan's, but no one has a majority. "There's no tiebreaker," says Al Wootten, an astronomer at the NRAO.

Yet Beasley doubts that the process would have been any smoother if Europe or the United States had taken the lead. For smaller projects, he says, with stakes on the order of millions of dollars, minority funding partners might accept some decisions that run counter to their interests. But with a project the size of ALMA, in which even a minority stake is hundreds of millions of dollars, funding agencies will fight to protect their interests. "No one is going to lose any significant decisions at that point," he says.

Beasley says that it would be better to create a strong central authority at the outset, and persuade funding agencies to grant it budgetary powers. There are precedents, particularly the treaty-governed European research institutes such as CERN, a particle-physics facility in Geneva, Switzerland, and the ESO itself, whose member states pay dues each year. And in 2011, the Square Kilometre Array created the SKA Organisation — a non-profit company based at the Jodrell Bank Observatory near Manchester, UK — which might give it the authority missing from the JAO.

But it is hard to imagine a funding agency such as the NSF — which answers to the US Congress — ceding control. So in the near term, big astronomy is likely to be governed by loose confederations, and the success of future mega-projects will depend on the savvy and sweat of the people within. Anyone who has served as ALMA director would know something about this. Each charmed rather than shouted his way to success. De Graauw had his courtliness; Massimo Tarenghi, director from 2003 to 2008, a certain puckishness. Cox's weapon might be positivity: the new director seems always to be grinning. Tarenghi hopes that those smiles will stay after Cox takes the helm in April. "The person that suffers most is the poor director," he says wryly.

COMING DOWN THE MOUNTAIN

By the end of the one-hour tour of ALMA, Pierre Cox is in fact suffering — from oxygen deprivation. Yet he still seems to be on cloud nine. "I'm infinitely grateful. I'm honoured. I'm thrilled," he says. "This is one of the coolest places I've ever been." He gets in the car for the downhill journey and slips an oxygen saturation meter over his finger. First it reads 70%, then 76%. Not good. The driver, who has already put oxygen tubes in his own nose, calmly hands Cox a pressurized can of oxygen. Cox takes a squirt in his mouth and checks his numbers again. More than 90%. Much better.

Maybe it's the rush of oxygen to the brain, but Cox becomes an enthusiastic chatterbox. The high-redshift Universe will be just the beginning, he says. He won't be completely satisfied following up on the objects others have already spotted. "There will always be surprises," he says. The car passes a vicuña (a relative of the llama) standing sentinel at the lip of a gully. ALMA's dishes have vanished behind the edge of the plateau. The OSF appears in the distance below, white rooftops shimmering as the desert heats up for the day. By 4,000 metres, the air is getting thicker. The oxygen has a soporific effect. De Graauw's head begins to nod. Cox yawns loudly. Inexorably, his eyelids close.

Behind him, on an isolated plain at the top of the world, the eyes of ALMA remain open, alert to the earliest glimmers of the Universe. ■

Eric Hand *is US news editor for* Nature.

© 2013 Macmillan Publishers Limited. All rights reserved



DALE EDWIN MURRAY

Campus 2.0

Massive open online courses are transforming higher education — and providing fodder for scientific research.

BY M. MITCHELL WALDROP

When campus president Wallace Loh walked into Juan Uriagereka's office last August, he got right to the point. "We need courses for this thing — yesterday!"

Uriagereka, associate provost for faculty affairs at the University of Maryland in College Park, knew exactly what his boss meant. Campus administrators around the world had been buzzing for months about massive open online courses, or MOOCs: Internet-based teaching programmes designed to handle thousands of students simultaneously, in part using the tactics of social-networking websites. To

© 2013 Macmillan Publishers Limited. All rights reserved

supplement video lectures, much of the learning comes from online comments, questions and discussions. Participants even mark one another's tests.

MOOCs had exploded into the academic consciousness in summer 2011, when a free artificial-intelligence course offered by Stanford University in California attracted 160,000 students from around the world — 23,000 of whom finished it. Now, Coursera in Mountain View, California — one of the three researcher-led start-up companies actively developing MOOCs — was inviting the University of Maryland to submit up to five courses for broadcast on its software platform. Loh wanted in. "He was very clear," says Uriagereka. "We needed to be a part of this."

Similar conversations have been taking place at major universities around the world, as dozens — 74, at the last count — rush to sign up. Science, engineering and technology courses have been in the vanguard of the movement, but offerings in management, humanities and the arts are growing in popularity (see 'MOOCs rising'). "In 25 years of observing higher education, I've never seen anything move this fast," says Mitchell Stevens, a sociologist at Stanford and one of the leaders of an ongoing, campus-wide discussion series known as Education's Digital Future.

The ferment is attributable in part to MOOCs hitting at exactly the right time. Bricks-and-mortar campuses are unlikely to keep up with the demand for advanced education: according to one widely quoted calculation, the world would have to construct more than four new 30,000-student universities per week to accommodate the children who will reach enrolment age by 2025 (see go.nature.com/mjuzhu), let alone the millions of adults looking for further education or career training. Colleges and universities are also under tremendous financial pressure, especially in the United States, where rocketing tuition fees and ever-expanding student debt have resulted in a backlash from politicians, parents and students demanding to know what their money is going towards.

When MOOCs came along, says Chris Dede, who studies educational technologies at Harvard University in Cambridge, Massachusetts, they promised to solve these problems by radically expanding the reach of existing campuses while streamlining the workload for educators — and universities seized on them as the next big thing.

There is reason to hope that this is a positive development, says Roy Pea, who heads a Stanford centre that studies how people use technology. MOOCs, which have incorporated decades of research on how students learn best, could free faculty members from the drudgery of repetitive introductory lectures. What's more, they can record online students' every mouse click, an ability that promises to transform education research by generating data that could improve teaching in the future. "We can have microanalytics on every paper, every test, right down to what media each student prefers," says Pea.

MOOC companies still face challenges, such as dealing with low course-completion rates and proving that they can make profit. And they have a lot of convincing to do among faculty members, says Uriagereka. "Some salivate and can't wait to be a part of it," he says, noting that his university had 20 volunteers for its 5 inaugural MOOCs. "Others say, 'Wait a minute. How do we preserve quality? How do we connect with students?'"

LARGE-SCALE PEDAGOGY

MOOCs are largely a product of one corridor in the Stanford computer-science department, where the offices of Andrew Ng, Daphne Koller and Sebastian Thrun are just a few steps apart. But they are also the fruit of research dating back to at least the 1990s, when the explosive worldwide growth of the Internet inspired a multitude of efforts to exploit it for education. Campus administrators tended to regard such projects as a sideshow — the higher-education financial crunch was not quite as serious back then — so most experiments were the work of committed individuals, departments or research centres. But with the relentless advance of technologies such as broadband, social networking and smart phones, researchers' interest continued to grow.

Ng got involved in 2007 because he wanted to bring Stanford-quality teaching to "the people who would never be able to come to Stanford", he says. Following a path blazed by the open-source software movement, and by earlier open-source education initiatives, he started a project to post online free lecture videos and handouts for ten of Stanford's most popular engineering courses. His approach was fairly crude, he admits: just record the lectures, put them online and hope for the best. But to his astonishment, strangers started coming up to him and saying, "Are you Professor Ng? I've been taking machine learning with you!" He began to grasp how far online courses could reach, and started working on a scaled-up version of his system. "When one professor can teach 50,000 people," he says, "it alters the economics of education."

> "When one professor can teach 50,000 people it alters the economics of education."

One of the many people he talked to about his work was Koller, who began developing her own online-education system in 2009. Whereas Ng looked outwards, Koller wanted to look inwards and reform Stanford's teaching on-campus. She particularly wanted to promote 'flipping', a decade-old innovation in which students listen to lectures at home and do their 'homework' in class with their teachers, focusing on the most difficult aspects or discussing a concept's wider implications. This lets the instructors concentrate on the parts of teaching most of them enjoy — interacting with the students — and relieves them of the repetitive lecturing that they often dislike.

Koller also wanted to incorporate insights from the many studies showing that passively listening to a lecture is a terrible way to learn (F. I. M. Craik and R. S. Lockhart *J. Verb. Learn. Verb. Behav.* **11,** 671–684; 1972). Following an approach pioneered by other online developers over the previous decade, Koller broke each video into 8–10-minute segments separated by pauses in which students have to answer questions or solve a problem. The idea was to get them to think about what they had learned; the deeper their engagement, studies showed, the better their retention.

Finally, to encourage greater interaction among the students themselves, Koller took a cue from social-networking sites such as Facebook and gave her system an online discussion forum. As Ng explains, the idea was to extend what happens in a face-to-face study group: "Students sit with their best friends, they work on problems together, they critique each others' solutions — lots of pedagogical studies show that these more interactive modes of student engagement result in better student learning."

Koller and Ng eventually realized that they could achieve both their goals — outreach and on-campus reform — by pooling their efforts. In late 2010, they started work on a software platform that would support discussion forums, video feeds and all the other basic services of an online course, so that an instructor only had to provide the content. But making social interaction work on a large scale turned out to be a research project of its own, says Ng. For example, standard online discussion forums are a fine way to bring communities together — for 100 or so users. "With 100,000 it gets more complicated," he says. Hundreds of students might end up asking the same question. So the developers implemented a real-time search algorithm that would display related questions and potential answers before a student could finish typing. Ng and Koller also let students vote items up or down, much like on the link-sharing website Reddit, so that the most insightful questions would rise to the top rather than being lost in the chatter.

The two researchers even set the system up so that students could mark one another's homework for essay questions, which computers

© 2013 Macmillan Publishers Limited. All rights reserved

can't yet handle. Not only is such a system essential to scaling up learning, says Koller, but it also turns out to be a valuable learning experience. And experiments have shown that if the criteria are spelled out clearly, grades given by the students correlate strongly with those given by the teacher (R. Robinson *Am. Biol. Teach.* **63,** 474–480; 2001).

By early 2011, Ng and Koller were planning to demonstrate the platform on campus, and other faculty members were paying attention. Among them was Thrun, a robotics researcher who was splitting his time between Stanford and Google in Mountain View, where he worked on the development of driverless cars.

It was Thrun's idea to go big, using a platform of his own based in part on Ng and Koller's ideas. He says that he was scheduled to teach an artificial-intelligence course that autumn, along with Peter Norvig, Google's director of research, "and I thought it was a social responsibility to take it online, so we could reach more than the 200 students we would get at Stanford". But even he hadn't imagined how big it would get. This was the course that registered 160,000 people from 195 countries after just one public announcement, a post to an artificial-intelligence mailing list. "It shocked everybody," he says.

In response, Ng took Koller's machine-learning course public using their platform, while department chair Jennifer Widom did the same with a database course. Each attracted roughly 60,000 students. With those numbers, venture-capital funding quickly followed.

Thrun announced his company Udacity in January 2012. Arguing that most professors don't have a clue about how to exploit the online medium, he and his colleagues elected to develop their courses in-house, working with education experts to make the pedagogy as effective as possible.

Ng and Koller announced Coursera in April 2012, and took the opposite tack. They partnered with big-name universities — Stanford and three others, to start — and let them provide the content while Coursera provided the hosting and software platform.

Anant Agarwal, former head of the computer science and artificial-intelligence laboratory at MIT, had been experimenting with online learning for a decade, developing an electric-circuit simulation package called WebSim that tried to give online students an effective substitute for hands-on laboratory experience. In December 2011, inspired by goings on at Stanford, he launched MITx: an independent, not-for-profit company that would offer massive online courses from MIT on an open-source basis. It became edX in May 2012, when Harvard joined.

At the same time, the term MOOCs, which had been circulating quietly in educational circles since it was coined in 2008, took off. Media accounts boomed, and company principals were soon giving talks at the popular Technology, Entertainment and Design (TED) conferences and the annual meeting of the World Economic Forum in Davos, Switzerland. As Koller told one interviewer: "I can't believe my life!"

270,000

Students enrolled in Udacity's Introduction to Computer Science MOOC

200,000

US university first-years intending to major in computer science

LEARNING CURVE

The MOOC companies can point to plenty of success stories. For example, the 7,200 students who completed Agarwal's electric-circuits MOOC in spring 2012 included an 81-year-old man, a single mother with two children, and a 15-year-old prodigy from Mongolia who got a perfect score on the final exam. Udacity's Introduction to Computer Science MOOC, currently its most popular, has enrolled more than 270,000 students.

But MOOCs have also had some teething problems. "Many people have no idea what they're in for when they commit to put a course online," says John Mitchell, a computer scientist and Stanford's first vice-provost of online learning. "Restructuring even one lecture into short, self-contained segments takes a fair amount of thinking." So does coming up with good, compelling questions to engage the students between the segments. Then there is the push for high-quality production, he says. "It takes many hours to produce one hour of quality video."

More worrisome are the MOOCs' dismal completion rates, which rarely rise above 15%. Completion has been a problem for distance learning ever since the first correspondence courses in the nineteenth century, says Dede. Only a small fraction of students have the drive and the perseverance to learn on their own, he says, and most people need help: "social support from their fellow students to help them keep going, and intellectual support from their professors and fellow students to help them figure out the material". At the moment, says Dede, the MOOC companies' peer-to-peer communication tools don't do nearly enough to provide that kind of help. "They're just kind of hoping that people will figure out from the bottom up how to support each other," he says.

The companies acknowledge that completion rates are a concern and that their platforms are still works in progress. "My aspiration isn't to reach the 1% of the world that is self-motivating," says Thrun, "it's to reach the other 99%." The companies are already working on enhanced social tools such as live video and text chat, for example.

And to observers such as David Krakauer, that is as it should be. "There are two ways to make something new," says Krakauer, a biologist who directs the Institute for Discovery at the University of Wisconsin–Madison. "You can design something that's perfect on paper, and then try to build it. Or you can start with a system that's rubbish, experiment and build a better one with feedback. That's the Silicon Valley style — but it's also the scientific way."

SILICON VALLEY STYLE

A Silicon Valley sensibility permeates the three big MOOC firms. For example, they all subscribe to the open-source ideal. "Charging for content would be a tragedy," says Ng. But they also see plenty of opportunities to make money using the 'freemium' model followed by Google and many other technology companies: give away the basic product to draw users, and then charge for premium add-ons.

One obvious add-on might be certification, says Ng. "You would get a certificate that verifies you took the course for a small fee like US$10–$30" — a potentially substantial revenue stream when enrolments are in six figures. In the future, the companies might also offer full university course credits for a fee; they are already working with accreditation agencies to arrange that.

Other possibilities include profiting from in-course mentoring services, career counselling — and charging universities for licensing. In October 2012, for example, edX licensed a circuit-theory MOOC designed by Agarwal to San Jose State University in California, where it was used as the online component of a flipped classroom experience. In return for the licensing fee, "the professors can offer the course on campus, tweak the course however they please, get access to students' grades and online activity, and all the analytics a teacher would want to see", says Agarwal. In this particular experiment, he adds, the San Jose course's usual 40% failure rate fell to 9%.

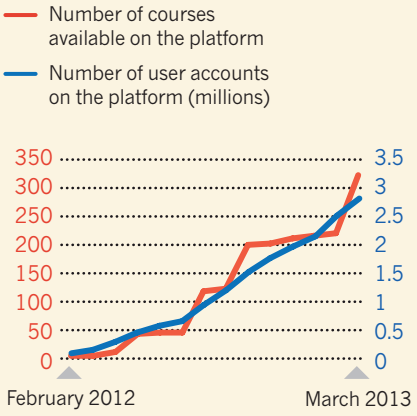
Analytics are another example of the Silicon Valley style, potentially allowing the MOOC companies to do for education what Internet giants such as Google or Amazon have done for marketing. In Coursera's case, says Koller, the platform monitors the students' every mouse click — "quiz submissions, forum posts, when and where a student pauses a lecture video, or rewinds, or moves to 1.5 speed".

© 2013 Macmillan Publishers Limited. All rights reserved

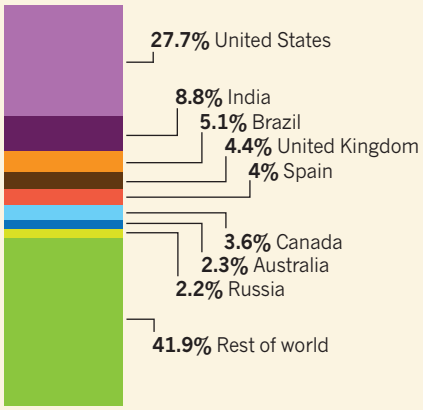
MOOCs rising

Over little more than a year, Coursera in Mountain View, California — the largest of three companies developing and hosting massive open online courses (MOOCs) — has introduced 328 different courses from 62 universities in 17 countries (left). The platform's 2.9 million registered users come from more than 220 countries (centre). And courses span subjects as diverse as pre-calculus, equine nutrition and introductory jazz improvisation (right).

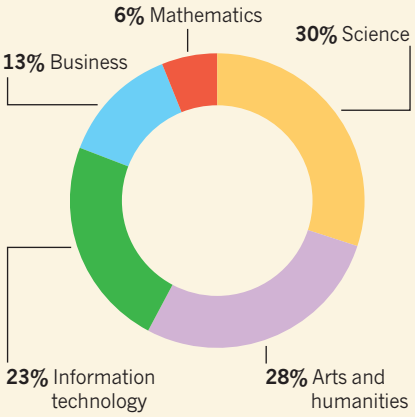
Supply and demand



Student origins



Courses offered



The company is constantly using these data as feedback, says Koller, both for refining the platform's user interface and for improving the course content. If 90% of the students start stumbling over the review exercises for a certain lecture, for example, then maybe it is time to revise that lecture.

"But anything we do is just the tip of the iceberg," says Koller. When data from individual students are multiplied by tens or hundreds of thousands of students per course, they reach a scale big enough to launch a whole new field of learning informatics — "big-data science for education", Pea calls it.

Learning informatics could provide an unprecedented level of feedback for colleges and universities, says Stevens: "We haven't measured learning in higher education very often, very consistently or very well — ever." Academics have endlessly studied factors that are associated with university enrolment and success, such as race, parental income and school achievement. They have also studied what happens after graduation: the higher earnings and other benefits that college confers, on average, over a lifetime.

"What we don't know is how college performs this magic," says Stevens. "We certainly don't know the extent to which digitally mediated college experiences will deliver the same returns as a four-year residential experience." Now, however, he and his colleagues can begin to see what education science will look like as it merges with data analytics. Instead of looking at aggregate data about students on average, for example, researchers can finally — with appropriate permissions and privacy safeguards — follow individual students throughout their university careers, measuring exactly how specific experiences and interactions affect their learning. "It's thrilling," he says, "a huge intellectual frontier."

What remains to be seen is how higher education will change in response to the new technology. Maybe not much, says Dede. Yes, the major universities will extend their courses beyond their own campuses; the MOOCs have already shown them that they can do so with relatively little effort and potentially large profits. But the MOOC founders' other goal — fundamental reform in on-campus teaching — is a much tougher proposition.

"Universities think of themselves as being in the university business, not the learning business," explains Dede. That is, they mostly take their existing structures and practices as given, and look to MOOCs and other online technologies as a way to do things more cheaply. But experience with earlier innovations such as personal computing shows the limits of that approach, he says: real gains in the productivity and effectiveness of learning will not come until universities radically reshape those structures and practices to take full advantage of the technology.

No one knows exactly where that restructuring might end up. Lectures becoming a rarity, for example? Vast numbers of students getting their degrees entirely online? But the revolution has already begun, says Stevens. Major universities such as Stanford are taking the lead, "trying to integrate and embed digital learning into the fabric of the entire university" — and trying to master the new technology before it masters them.

Virtually everyone participating in this upheaval agrees on one thing. Colleges and universities will change — perhaps dramatically — but they will not disappear. "No one says that all education has to be online," says Thrun. "Sometimes, a classroom is better." Especially in communal endeavours such as science, "education is more than just knowledge", says Dede. "It's abilities like leadership and collaboration, and traits like tenacity", all of which are best learned face to face.

An unspoken irony weaves through almost every discussion about MOOCs: thanks to innovations such as flipping, online technology's most profound effect on education may be to make human interaction more important than ever. As Krakauer puts it, "what's absolutely clear is that the very large lecture hall can be completely replaced: there's no value added over watching it at home on an iPad screen with a cup of tea. But there is also no substitute for a conversation." ■

M. Mitchell Waldrop *is a features editor for* Nature *in Washington DC.*

© 2013 Macmillan Publishers Limited. All rights reserved

COMMENT

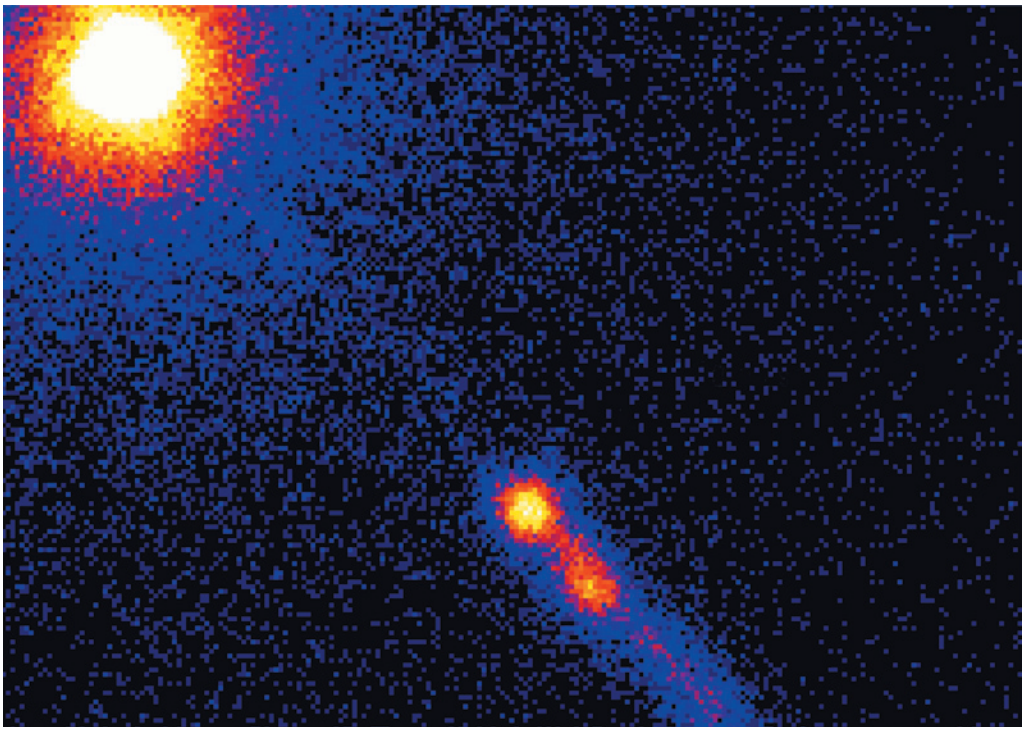
COMMUNITY A call for an end to scientific snobbery about where people work **p.167**

HISTORY Cholera cartographer and father of epidemiology, John Snow, remembered **p.169**

BOTANY A biography of the gingko, the tree that has outlasted the dinosaurs **p.170**



CULTURE Stunning ice-age art exhibition disappoints through lack of science **p.173**



NASA/CXC/SAO/H.MARSHALL *ET AL.*

X-ray observations show that quasar 3C273 shoots out a jet of plasma blobs that seem to move faster than light.

Quasars still defy explanation

Fifty years after finding that these cosmic beacons lie far away, astronomers need to think harder about how they radiate so much energy, says **Robert Antonucci**.

Arthur C. Clarke posed one explanation for why no extraterrestrial life forms have been in touch in his 1953 novel, *Childhood's End*. The book describes a Galactic club of advanced civilizations that have a policy not to interfere in cultures at a primitive stage of evolution, such as our own. But once a society masters nuclear weapons and interstellar travel and becomes dangerous, the Galactic authorities introduce themselves and their rules, which include a ban on wars.

Astronomy's childhood ended 50 years ago with a discovery that made us full citizens of the Universe. In 1963, the first measurement of the distance to a quasar — a radio source that looks like a star in visible light — showed it to be an enormously powerful beacon lying billions of light years away[1]. Until then, astronomy had been limited to exploring our local patch of space time, in which everything looks familiar. Before quasars, the distant Universe was tantalizingly out of reach.

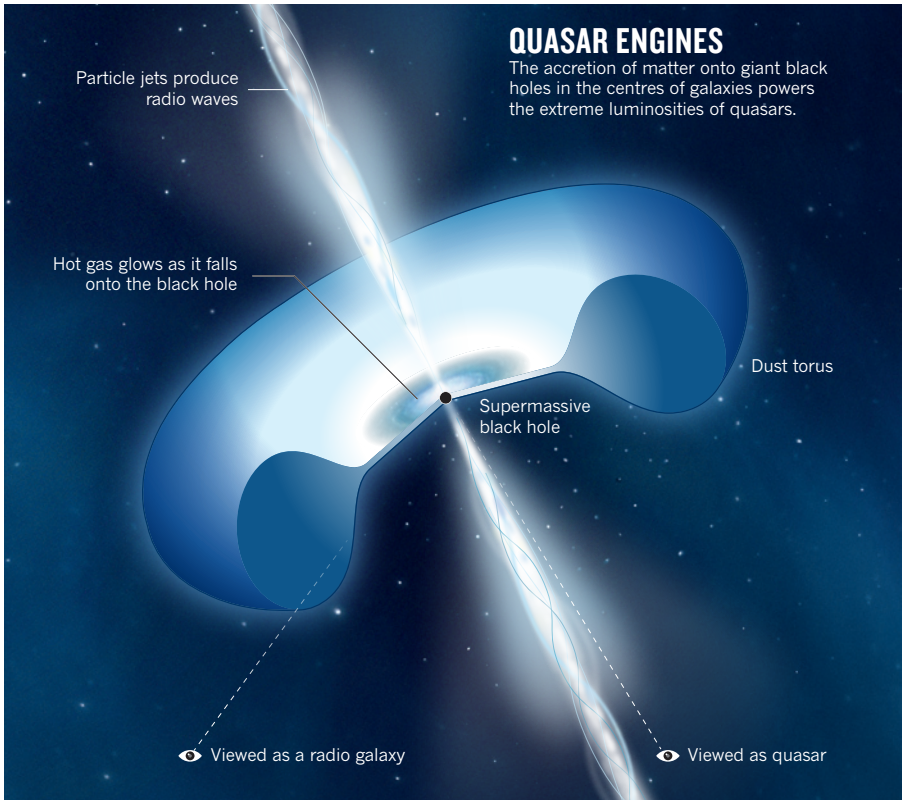
Quasars are immensely bright. From the central point in a galaxy, they emit as much energy as thousands of giant galaxies from a region as tiny as the Solar System. They radiate energy across the electromagnetic spectrum, from radio waves to γ-rays. Many expel jets of particles at near-light speed, which inflate vast particle clouds or 'lobes' that measure millions of light years across and emit radio waves.

Light that has travelled from distant quasars offers us a glimpse back in time. Since the discovery in the 1920s that the Universe is expanding, cosmologists have known that the cosmos has a finite age of about 13.7 billion years. Astronomers have ▶

© 2013 Macmillan Publishers Limited. All rights reserved

had great fun detecting distant quasars, pushing ever closer to the Big Bang as technology has improved. But they have yet to understand the detailed physics of how quasars emit such enormous amounts of energy.

There is a consensus that quasars and other active galactic nuclei (AGN) are powered by the accretion of gas and stars onto giant black holes in galaxy cores, but the details are still mysterious (see 'Quasar engines'). Many theoretical models of quasars have little or no predictive power, and so, in my opinion, are of little or no value. Quasar researchers must recognize this problem and focus on understanding the physics of black holes.



LOOKING BACK

The first objects known to lie far beyond the Milky Way were radio galaxies, whose lobes were picked up by radio telescopes in the 1950s. Using optical telescopes, their distances were measured from the 'redshifts' of spectral lines of elements such as hydrogen, stretched to longer wavelengths owing to the passage of light through the expanding Universe. By 1960, radio galaxy 3C295 (the 295th object listed in the third Cambridge catalogue of radio sources) was the champion at a redshift of 0.464, meaning its light departed 4.8 billion years ago. But radio galaxies are dim and hard to spot from far away.

The March 1963 announcement of the first redshift for a quasar made it possible to dramatically expand the cosmic horizon. In a pair of papers in *Nature*, Maarten Schmidt[1] showed that the bright source 3C273 — another object from the Cambridge catalogue — has a redshift of 0.158 and lies 2.4 billion light years (736 megaparsecs) away, and Bev Oke[2] revealed that its spectrum was unlike a normal star. The extreme power implied by its brightness over that distance was shocking. Much higher redshifts for other quasars soon followed.

Astronomers Fred Hoyle and William Fowler realized in early 1963 that the immense energies of the radio lobes could only be created by the gravitational collapse of a very massive object that was a hundred million times heavier than the Sun. In 1969, Donald Lynden-Bell argued in *Nature* that such a monster would be crushed down under its own weight to form a supermassive black hole. The collapse of large masses to tiny sizes, which means almost dividing by zero in the equation for gravitational potential energy, puts much of the zing into astronomy.

Some quasar jets, including one in 3C273, appear to shoot blobs of plasma out of their nuclei at ten times the speed of light — seemingly breaking a rule of relativity. In the 1970s and early 1980s, astronomers Peter Scheuer and, independently, Roger Blandford and Martin Rees explained these 'superluminal' motions of the blobs as illusions. These arise because time delays in the light emerging from bright spots in the jets produce fast-shifting patterns, in a kind of time-lapse photography.

In the 1980s and early 1990s, astronomers realized that several key properties of radio galaxies and quasars were consistent with the idea that these two classes differ only in orientation with respect to the line of sight. This led to 'unified models' that explain how characteristics of AGNs vary with jet angle[3].

Quasar jets that point nearly towards us show superluminal motions; misdirected ones do not. Radio galaxies have jets in the plane of the sky. Separately, it was shown in the 1980s using optical spectropolarimetry and other means that many AGNs apparently lacking the broad emission lines that characterize quasars do show them in polarized light, scattered towards us by clouds of gas[3]. The unified model has allowed astronomers to piece together common characteristics of different classes of AGN. But it has not told us much about the central black hole and how its energy emerges.

We now know that supermassive black holes are common to most galaxies. In the 1980s and 1990s, observations of fast motions of gases and stars hinted that nearby galaxies host inactive central black holes[4]. Since then, astronomers have realized that quasars mark a phase in the life of galaxies, when their central black holes are lit up by accreting matter. This trait was more common in the past, so there are fewer quasars today. Now starved of fuel, black holes linger in galaxies, including our Milky Way.

LITTLE PROGRESS

Have we made good progress in understanding quasars in five decades? I do not think so. The theory of radio sources has not changed significantly in the past 30 years[5]. Basic questions remain: do the jets and lobes comprise electrons and protons or electron–positron pairs? Do the protons carry a lot of energy, as cosmic rays do? Is the energy divided evenly between electric and magnetic fields? Without answers to these, we can set only lower limits on how much energy the jets and lobes hold.

In my opinion, the greatest limiting factor in understanding quasars is not a lack of intelligence, effort or creativity, nor is it a dearth of fantastic new facilities. It is a widespread lack of critical thought among many researchers. Theories are being published that have already been ruled out by observations. Observers cling to falsified theories when interpreting their data. A vast amount of work has been wasted.

Most of the AGN community is mesmerized by unphysical models that have no predictive power. Smart people continue to use and refine overly simple versions of 'accretion disk' models of matter spiralling towards the black hole, including disproved assumptions such as that the disk structure changes only slowly[6,7]. These models simply don't match the observations without lots of special pleading. The properties of small accretion disks that are inferred to exist around stellar-mass black holes cannot be scaled up to explain the spectra of much more luminous quasars. Models and observations of quasar

© 2013 Macmillan Publishers Limited. All rights reserved

emissions violate even basic rules such as the Stefan–Boltzmann law, which states that the total energy radiated by a black body scales with its temperature by the fourth power. Recent work highlights other problems, such as that much of the optical light from quasars comes from regions that are much larger than disk models predict[8].

One revealing faux pas highlights the fact that some astronomers like to 'see what they believe'. In 1984, many in the field rejoiced when the first sensitive orbiting ultraviolet telescope obtained a temperature for one AGN that agreed with the prediction of the accretion disk model. Then, an amateur astronomer spotted a missing factor of ten in Newton's gravitational constant in the analysis. The model no longer fitted and the authors quickly issued a correction. Yet the erratum[9] is hardly cited, whereas the original paper has more than 100 citations.

The AGN field seemed to see a breakthrough in 1995, when orbiting X-ray telescopes became sensitive enough to detect high-energy iron emission lines that are broadened by rapid motions and gravity close to the black hole. Astronomers have again struggled to find a sensible theory. The prevailing scenario is that the lines are produced by a mysterious X-ray source that hovers above a disk of cold material. Although there have been preliminary claims of validation of this theory[10,11], in my view there is no good evidence yet that the iron line strengths follow variations in the X-ray continuum brightness closely and with little delay as one would expect.

The X-ray evidence for the disk reprocessing model is minuscule compared with more developed studies of optical line variability. Even stranger, bumps on the iron emission-line profiles persist for days, implying that the X-ray source somehow manages to float above the same spot even while the disk spins, which seems impossible. Adjustments to the theory, such as that the bending of light scrambles the responses of the lines to the continuum, only add more complexity. X-ray and ultraviolet AGN astronomers rarely talk to one another, even though both groups work on processes in the inner disk.

I see fewer and fewer serious theory papers on AGNs, but there is a burgeoning effort to find ever more quasars in surveys. It is as if, having given up on understanding AGNs, the community now focuses on the more modest goal of counting them. Sadly, the statistical analysis of astronomical surveys leads to further problems, such as claiming causal links in plots of dependent variables[12].

WHERE NEXT?

We have found thousands of quasars in the past 50 years, but we still don't have good physical models for how they radiate their prodigious energy. Without predictive theories we have nothing — our best hope for understanding quasars is that extraterrestrials might drop in and explain them to us.

Astronomers working on AGNs need to take a cold, hard look at their models and discard those that are unrealistic or have been falsified by observations. We need to use more sensitive X-ray telescopes to make detailed studies of the emission lines arising in the vicinity of the black hole. And we need advanced computational modelling of black holes, including magnetism, fluid dynamics, thermal balance and radiation. I hope to do my bit by elucidating the raw spectrum of radiation produced by the central engines of active galaxies, after removing distortions caused by surrounding gas and dust[13].

I urge my junior colleagues to spend 15 minutes a day thinking, palms down and eyes on the ceiling. That's just 3% of their time. As I saw recently on a Californian bumper sticker: "Don't just do something, sit there." ■

Robert Antonucci *is professor of astrophysics at the University of California, Santa Barbara, California 93106, USA.*
e-mail: skiantonucci0@gmail.com

1. Schmidt, M. *Nature* **197,** 1040 (1963).
2. Oke, J. B. *Nature* **197,** 1040–1041 (1963).
3. Antonucci, R. *Annu. Rev. Astron. Astrophys.* **31,** 473–521 (1993).
4. Kormendy, J. & Richstone, D. *Annu. Rev. Astron. Astrophys.* **33,** 581–624 (1995).
5. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Rev. Modern Phys.* **56,** 255–351 (1984).
6. Blaes, O. in *The Central Engine of Active Galactic Nuclei* (eds Ho, L. C. & Wang, J.-M.) Vol. 373, 75 (ASP Conference Series, 2007).
7. Alloin, D. *et al. Astrophys. J.* **308,** 23–35 (1986).
8. Dexter, D. & Agol, A. *Astrophys. J. Lett.* **727,** L24 (2011).
9. Ulrich, M. H. *et al. Mon. Not. R. Astron. Soc.* **209,** 479 (1984).
10. Zoghbi, A. *et al. Mon. Not. R. Astron. Soc.* **422,** 129–134 (2012).
11. Zoghbi, A. *et al.* Preprint at http://arxiv.org/astro-ph1302.1761 (2013).
12. Antonucci, R. *Astron. Astrophys. Trans.* **27,** 557–602 (2013).
13. Kishimoto, M. *et al. Nature* **454,** 492–494 (2008).

Scientists are snobs

It is a mistake to dismiss the people and projects coming out of lesser-known institutions, argues **Keith Weaver** — they have strengths too.

We all do it. Pressed for time at a meeting, you can only scan the presented abstracts and make snap judgements about what you are going to see. Ideally, these judgements would be based purely on what material is of most scientific interest to you. Instead, we often use other criteria, such as the name of the researchers presenting or their institution. I do it too, passing over abstracts that are more relevant to my work in favour of studies from star universities such as Stanford in California or Harvard in Massachusetts because I assume that these places produce the 'best' science.

As someone who is based at a less well-known institution, the University of South Dakota in Vermillion, I see other scientists doing the same to me and my students. In many cases, this is a loss: to my students and their projects, which could have benefited from the input, and to the investigators who might have missed information that could have been useful in their own work.

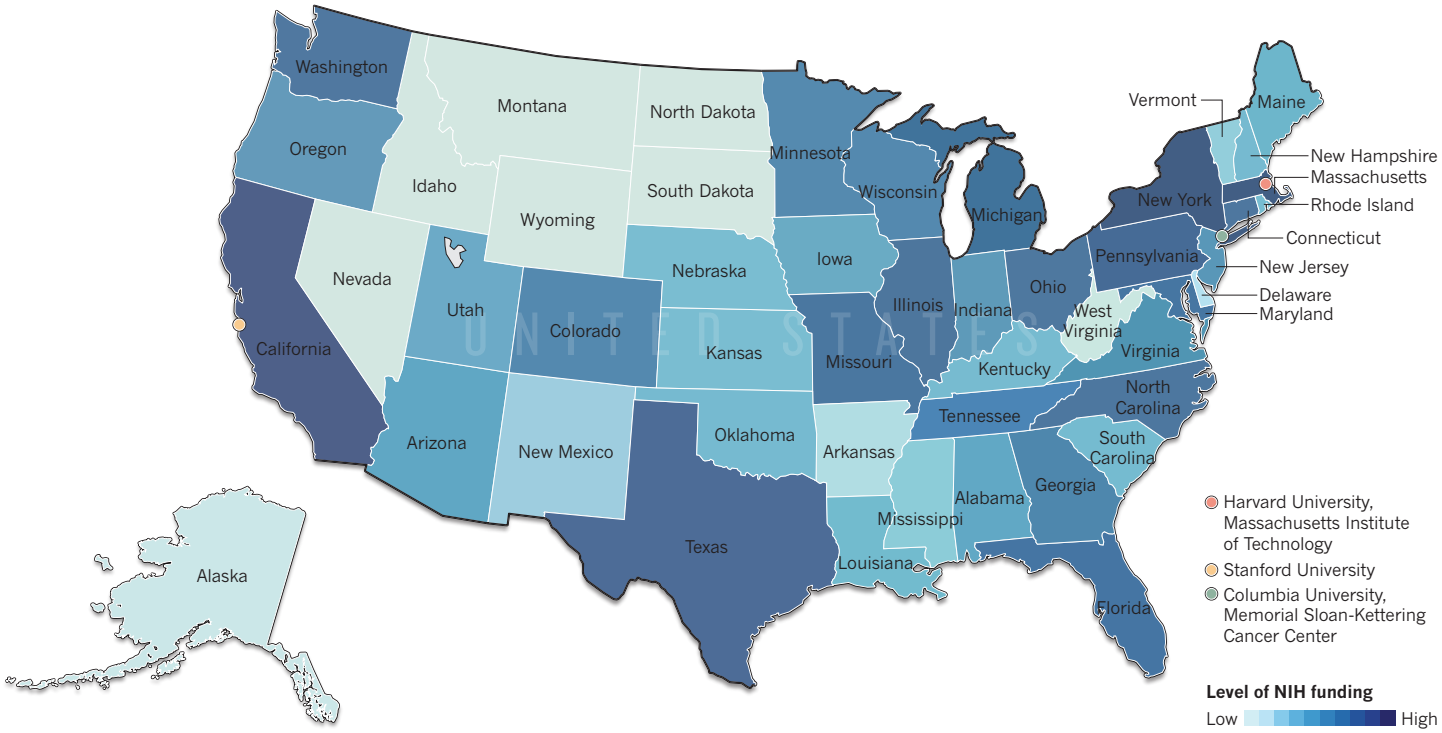
Of potentially greater impact is the effect this scientific snobbery has on citation practices. My laboratory was the first to describe a system in a Gram-positive bacterium that uses a small, non-coding RNA to block expression of a toxin in plasmid-containing cells, ensuring that only those cells survive (K. E. Weaver *RNA Biol.* **9,** 1498–1503; 2012). The system allows the persistence of plasmids that carry antibiotic-resistance genes, and points to a new mechanism for subverting that resistance. If we can find a way to interfere with the function of the regulatory RNA, we could conceivably induce the cells to commit suicide. Follow-up work by colleagues in larger labs has demonstrated that hundreds of related systems exist on plasmids, phage and chromosomes in numerous Gram-positive bacteria, including some important pathogens.

Although these colleagues were diligent in citing my original work, recent papers describing a system similar to my own in another bacterium cited only a more recent study from a large US National Institutes of Health laboratory. Losing this and other worthy citations could ultimately affect my ability to get promoted and attain grants. ▶

© 2013 Macmillan Publishers Limited. All rights reserved

MIDDLING WORK

Academics based in states such as those that persistently receive less funding from the US National Institutes of Health (NIH) can be snubbed by other researchers, regardless of the quality of their science.



SOURCE: NIH

Such snobbery arises from a preconceived idea that many scientists have — that people end up at smaller institutions because their science has less impact or is of lower quality than that from larger places. But many scientists choose smaller institutions for quality-of-life reasons, as I did. I moved to the University of South Dakota in 1989 after starting a family. Having lived in several large cities during my training, I was attracted to the small-town atmosphere of Vermillion with its good schools and safe streets. The smaller size of the medical school class and the graduate programme also made for a more personal learning environment. Although I was expected to publish papers and obtain grants, it was clear that I would not have to receive multiple grants to pay most of my salary, get promotions and obtain a tenured post.

I am not blind to the disadvantages. It is more difficult for me to get graduate students and postdocs to join my lab, because of the belief that working at a bigger department is better for their career. And even if the size of the institute isn't an issue, many people prefer to live in a city rather than spend several years in a small town. As a result, my lab typically contains only a handful of people and so has to select research questions that are narrower in scope.

This focus brings advantages. With fewer people to supervise, when my collaborator wants a gel done, I do it myself. This keeps me grounded in the realities of experimental science, and means that my problem solving and interpretive skills stay sharp. It also provides a constant reminder of why I got into this job in the first place.

My students speak to me nearly every day, and do not have to rely on fleeting sightings of me on campus or in a monthly meeting, as is the case in many larger departments. I feel confident about the work that they are doing and the quality of their data, because I am involved every step of the way. So although the volume and scope of work I produce might be less than I could accomplish at a larger institution, I like to think that the quality is comparable. I believe that this makes the students' experience more enriching — the best and brightest students might not always apply to my lab, but those who do have more of a chance to improve their education.

"Many scientists choose smaller institutions for quality-of-life reasons."

A small place is also less bureaucratic. Several years ago, our medical school did away with the traditional department structure and formed a single Division of Basic Biomedical Sciences. This reorganization would have been much more difficult at a larger department with more faculty members. The greater interaction between disciplines has benefited academic staff and students alike. Group lab meetings include people working in other fields, such as anatomy or eukaryotic biology, and have led to interesting collaborations. I also feel that I have more of an impact on administrators than I would at a top-tier institution — I know the dean well, and can walk into his office and offer my opinion.

Unfortunately, these traditional academic values might not always pay off. As far back as 1965 — when biomedical research was a far less cutthroat business — one study showed that scientists gained more recognition from a name tag that mentioned a prestigious institution than from their own highly productive work (D. Crane *Am. Sociol. Rev.* **30,** 699–714; 1965). And a map of funding from the National Institutes of Health makes it clear that many states are persistently underfunded (see 'Middling work). Whether that pale-blue gap in the middle of the country is the cause or the result of preconceptions about scientists from those regions is unclear. There are also fewer researchers in these empty spaces than in California or Massachusetts, but I am one of them.

At your next conference, try to stop at a poster that interests you from a smaller institution. I guarantee that the author will appreciate it. Talk to the scientists; visit their lab to see the quality of their science for yourself. Consider including more scientists from small places in your collaborations. We're all snobs sometimes, but awareness is the first step to changing behaviour. ■

Keith Weaver *is professor of microbiology at Sanford School of Medicine, University of South Dakota, Vermillion, South Dakota 57049, USA.*
e-mail: keith.weaver@usd.edu

© 2013 Macmillan Publishers Limited. All rights reserved



MARY EVANS PICTURE LIBRARY

London's polluted water supply was still the subject of satire in 1866.

IN RETROSPECT

On the Mode of Communication of Cholera

W. F. Bynum reassesses the work of John Snow, the Victorian 'cholera cartographer'.

Tomorrow is the bicentenary of the birth of John Snow, one of the founding fathers of epidemiology and anaesthesiology. His posthumous reputation is much greater than any he enjoyed during life.

Snow's epidemiological fame rests primarily on the second edition of his *On the Mode of Communication of Cholera* (1855), which rests, in turn, on his meticulous mapping of the disease in London and creative use of figures to show how cholera is transmitted. An ascetic man who never married, he died young; but his life was one of discipline, ambition and honest toil.

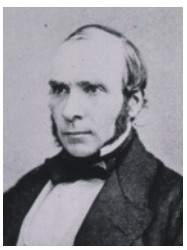
Born into a working-class family in York, UK, Snow was trained by apprenticeship to general practitioners and some courses in medical schools in north-east England, with little financial backing. He nevertheless based his career in London. He set up a general practice, worked in the outpatient's department at Charing Cross Hospital and became medical officer to several convalescent homes set up by working men's clubs.

London suffered three severe cholera epidemics during Snow's lifetime, in 1832, 1848–49 and 1853–54. In the first edition of *On the Mode of Communication of Cholera*, published in 1849, Snow extended arguments he had long held: that cholera is a specific disease spread by water contaminated by faeces. At the time it was generally thought to be a nonspecific fever transmitted through 'miasma' in the air. Far from claiming that his ideas were original, Snow scoured the literature and badgered acquaintances such as the epidemiologist William Farr for corroboration. He found some, but in many instances offered his own interpretations of the facts that they presented to him.

That foul water was the primary mode of cholera's spread was his mantra, and during the 1854 epidemic Snow conducted two studies to test it that would form the basis of his book's famous second edition. The better-known study focused on the Soho outbreak, in late August and early September, which centred around a pump in Broad (now Broadwick) Street. Some 600 people died. Snow used government death-registration data and house-to-house enquiries to map the victims' residences, showing their proximity to the pump. In a number of instances he was also able to confirm that victims had drunk water from the well, into which an open sewer drained. As in much of London, many houses in the locale were not yet connected to central water or sewage systems.

Snow convinced local officials to remove the pump handle, although the gesture was probably symbolic, because the epidemic was already subsiding. He further made his case by showing that several people who lived elsewhere in London had also drunk water from the well, and come down with cholera. Moreover, few residents at a local workhouse with its own well suffered, despite overcrowding. Employees of a nearby brewery were also essentially immune from the epidemic. One of the perks of their job was an allowance of free beer, so they didn't bother with water — a conclusion that Snow, one-time honorary secretary of the Medical Temperance Society, would have found ironic.

The second of Snow's epidemiological studies of 1854 may be referenced less often but was actually even more impressive. Back in 1848, Snow had incriminated two London water companies — one in Lambeth, one in Southwark and Vauxhall — with servicing their south London clients with contaminated water. Between the epidemics of that year and 1854, the Lambeth company switched its source from the Thames in central London to Thames Ditton, an upstream village some 12 miles away, and began to filter its water. In 1854, Snow began the hard task of tracing the contamination.



NIH/NLM

On the Mode of Communication of Cholera
JOHN SNOW
John Churchill : 1849, 1855 (2nd edition)

That second study involved some 300,000 people, from all occupations, social standing, ages and genders. Getting the relative death rates from cholera for each company was not easy: in many instances the people Snow questioned did not remember which company sold them water, and often houses in the same street were connected to one ▶

© 2013 Macmillan Publishers Limited. All rights reserved

or the other supplier almost at random. Snow's revelations encouraged Farr, who was also keeper of statistics at the General Register Office, to use his office to further the analysis. Both Snow and Farr found that householders drawing water from the Southwark and Vauxhall company were many times more likely to die from cholera.

Snow could not identify any specific causative agent, but he argued that the material cause of the disease acted like a living organism, because it could reproduce. Terming it "organised matter", Snow interpreted its incubation period as the time it took for the initial dose to reproduce sufficiently to cause actual disease. He never knew of the Italian microscopist Filippo Pacini's description of the comma-shaped bacillus (which causes cholera) during an 1854 outbreak of the disease in Florence. Snow had died by the time that germ theory became a viable proposition thanks to Louis Pasteur, Robert Koch (generally lauded for discovering the cholera bacillus) and many others.

Snow had another string to his bow: anaesthetics. From the late 1830s he had been concerned with issues surrounding asphyxia and artificial respiration in newborns, and so was an expert on the administration of gases to the lungs. He embraced the arrival of inhalation anaesthesia in 1846. Within a month of ether's first public use in Britain, in December of that year, Snow was researching its chemical properties, and he was administering it for surgeons by the following February.

His general practice soon centred on the administration of anaesthesia in leading London hospitals. He was even called in to give Queen Victoria 'blessed relief' during the births of her last two children. But his lasting contributions lay in the equipment he devised to control dosages of anaesthetic gases, and his research (often on himself) on ether and chloroform, which contributed to the more accurate description of the stages that patients pass through as the anaesthetic takes hold.

Snow was diagnosed with kidney disease more than a decade before his death, and reluctantly, on his doctors' advice, abandoned his vegetarian diet and his total abstinence from alcoholic beverages. One can surmise that his kidney disease contributed to raised blood pressure: he died six days after suffering a stroke, aged just 45.

His reputation then was higher as an epidemiologist than an anaesthetist, and routine filtration of water supplies was still in the future. Both contributions are highly valued now — you can raise a toast to Snow at the pub that bears his name, a stone's throw from the site of the Broad Street pump. ■

William Bynum *is professor emeritus of the history of medicine at University College London.*
e-mail: bynum2@me.com



Ginkgo trees can live for thousands of years.

BOTANY

A tree for all time

Sandra Knapp relishes a biography of the ginkgo, an arboreal survivor that has outlasted the dinosaurs.

Biographies seem to come in two sorts. The first is a dense compilation of deeds and misdeeds, dates and details of the dead; the second, a hagiography of celebrities still alive and kicking.

Ginkgo is neither. Its subject should by rights be dead, but having outlasted the dinosaurs, is still very much with us. This biography of the ginkgo tree offers a potent mix of science, history and culture, exploring how plants have changed our lives and our planet. And Peter Crane, a palaeobotanist by trade and all-round botanist by nature, is the perfect person to tell the tale.

Lucidly and accessibly, he takes us from the living plant, through its long and fascinating past as seen in the fossil record, and then returns us to the present in sections on the ginkgo in culture and the future of diversity. This tree, he shows us, can be seen as a metaphor for all life on Earth — seemingly fragile, but actually tough and likely to outlive *Homo sapiens*. With its meticulous footnotes, satisfying referencing and gripping narrative, I can see this book becoming a commuter's favourite for scientists and general readers alike.

Ginkgo biloba, as the maidenhair tree was named by the great Swedish botanist Carl Linnaeus, is one of a kind. It is a biodiversity loner, the only species in its genus, family and order. This was not always the case, as Crane shows. Ginkgo relatives first appear in the fossil record more than 200 million years ago, and the lineage has hung on through great extinctions. Long ago, *Ginkgo biloba* had many relatives, but they have all since died out, as it almost did itself.

As he traces these species' amazing journey, Crane introduces key concepts in evolution — the part extinction plays in producing current patterns of distribution, the fact that where things are not found is as important as where they are, and the role of contingency in creating today's diversity. I loved his equation of the stunning plant-fossil trove in Scotland's Rhynie Chert with the Burgess Shale of Canada, that phenomenal array of Cambrian life-forms made famous by Stephen Jay Gould's *Wonderful Life* (W. W. Norton, 1989). There is another book in waiting there.



Ginkgo: The Tree That Time Forgot
PETER CRANE
Yale University Press: 2013. 408 pp. $40

Crane's enthusiasm for fossil-hunting is very real. He recounts a stop in 1982 on a road in North Dakota, where he discovered some perfectly preserved ginkgo leaves and seed fossils from the Late Palaeocene (58.7 million to 55.8 million years ago). This extinct

DAVID STEPHENSON/LEXINGTON-HERALD-LEADER/MCT/GETTYIMAGES

© 2013 Macmillan Publishers Limited. All rights reserved

species is now known as *Ginkgo cranei*, but Crane is more taken with the joy of discovery than the honour of having a plant named for him.

As for the living tree, Crane fills us in on some of its peculiarities, not least its swimming sperm — studied by the reproductive-rights pioneer Marie Stopes during her career as a biologist — and the strange stalactite-like growth it uses for propagation. I went straight out to look at the ginkgo in front of London's Natural History Museum with new eyes.

Ginkgo ends with a beautiful section on the importance of this tree to human culture. It is revered for its longevity (the most venerable is some 3,500 years old) and beautiful form in China, Japan and Korea, and is one of the commonest street trees in the temperate zone in the West. The tree's loveliness is counterbalanced by a distinct autumnal odour: in Manhattan, where one-tenth of street trees are ginkgos, butyric acid in the seeds' fleshy covering gives off a whiff of rancid butter.

As Crane aptly puts it, its charisma — a mixture of unusual form, strange leaves and sheer staying power — has been an important element in its success and symbiosis with humans. Giant ginkgos are lovingly cared for and replaced when they start to die: the tree at London's Royal Botanic Gardens at Kew, planted some 250 years ago, is cared for as well as any national monument made by human hands.

The ginkgo was common over what is now the Northern Hemisphere between 100 million and 40 million years ago, but is elusive in the 'wild' today. Putatively wild ginkgos are known only from a couple of places in China, and they might be relics of old monasteries or human cultivation. Does it matter? On one level, no: that this species is still with us is itself to be treasured. On another, it does.

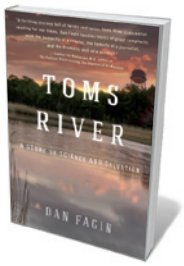
Earth's biodiversity has persisted through aeons, but no species has ever affected the planet as we have. Crane provides a wonderful precis of the politics of conservation, and how our own short ecological memories mean we often don't see the forest for the trees, however culturally important. People care for ginkgo trees because they are seen as special. But what of the rest of biodiversity? How can we conserve what we do not know?

The renowned naturalist Alfred Russel Wallace summed it up beautifully in 1863, saying that future generations "will charge us with having culpably allowed the destruction of some of those records of creation [species] which we had it in our power to preserve".

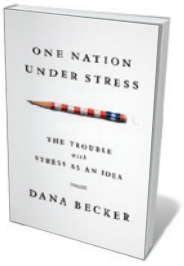
Ginkgo will inspire you to know and care for the organisms with which we share this planet in a new way. ■

Sandra Knapp *is a botanist in the Department of Life Sciences at the Natural History Museum in London.*
e-mail: s.knapp@nhm.ac.uk

Books in brief



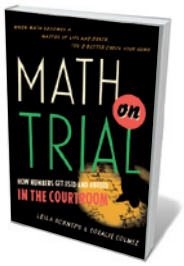
Toms River: A Story of Science and Salvation
Dan Fagin BANTAM *560 pp. $28 (2013)*
This hard-hitting account of cancer epidemiology in the New Jersey town of Toms River is a triumph. Hinging on a prolonged bout of toxic dumping by several companies up until the 1980s, journalist Dan Fagin's chronicle mixes reportage with science and industrial history. Thousands of drums of carcinogenic waste were buried in unlined pits, contaminating ground water; waste water was piped to coastal waters. In 2001, a landmark ruling linked some local cancers to local air and water pollution — a development with resonance, Fagin argues, for China's new industrial boom towns.



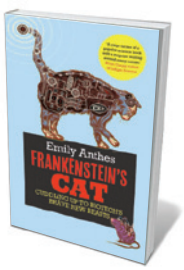
One Nation Under Stress: The Trouble with Stress as an Idea
Dana Becker OXFORD UNIVERSITY PRESS *256 pp. $35 (2013)*
Is stress a 'lifestyle problem', or the inevitable result of larger social and political inequities, imbalances and shifts? Sociologist Dana Becker argues that in the United States, the diffuse concept of stress now covers all kinds of tensions — effectively masking their triggers, from dual-career marriages to the frenetic, technology-driven pace of daily life. As a result, real social change in areas such as health care stalls. Becker's analysis tracks the evolution of 'stressism' from its origins as the 'price of progress', through medicalization, gender politics and conditions such as post-traumatic stress disorder.



Are We Being Watched? The Search for Life in the Cosmos
Paul Murdin THAMES AND HUDSON *224 pp. £16.95 (2013)*
Discoveries such as exoplanetary systems, water ice on Mars and extremophile bacteria on Earth, have energized the scientific quest for extraterrestrial life. Here, distinguished astronomer Paul Murdin, who discovered the first black hole in the Milky Way, uses findings from planetary and climate science, astronomy, the evolution of life on Earth and the missions of the Voyager and Galileo spacecraft as a springboard for astrobiological speculation. But this is a measured investigation: Murdin, ever clear-eyed, is well prepared to accept that we are unique.

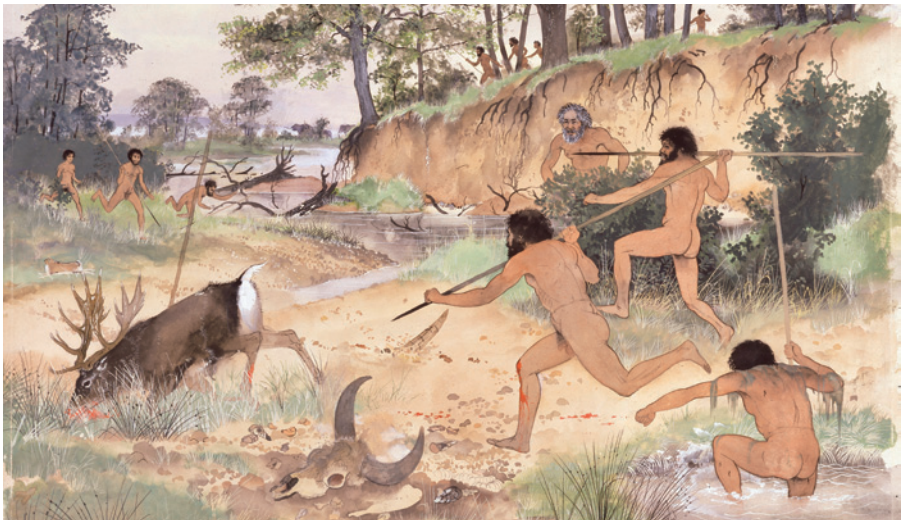


Math on Trial: How Numbers Get Used and Abused in the Courtroom
Leila Schneps and Coralie Colmez BASIC BOOKS *272 pp. $26.99 (2013)*
How does court evidence add up? Mathematicians Leila Schneps and Coralie Colmez show that it often doesn't: calculations, statistics and probability can be misused, with dire consequences. The authors forensically unpick ten cases in which justice was undone in this way, from recent trials to that of Alfred Dreyfus in nineteenth-century France. Dreyfus suffered a hellish imprisonment for treason after a handwriting expert exaggerated probability — a miscarriage of justice later exposed mathematically by Henri Poincaré.



Frankenstein's Cat: Cuddling Up to Biotech's Brave New Beasts
Emily Anthes ONEWORLD PUBLICATIONS *256 pp. £8.99 (2013)*
Tusked mice, transgenic goats lactating antithrombin, dogs with prosthetic testicles: science writer Emily Anthes reports from the wilder shores of animal biotechnology. A team at Fudan University in Shangai, China, for instance, has created 500 strains of modified lab mouse — and hopes to engineer 100,000. A Florida medical team has made a prosthetic tail for an injured dolphin. There is plenty more, but Anthes devotes the final word to bioethics, arguing that the advances are a chance to commit anew to animal well-being.

© 2013 Macmillan Publishers Limited. All rights reserved



NATURAL HISTORY MUSEUM, LONDON/SPL

A recreation of a Neanderthal hunting party in prehistoric Britain.

EVOLUTIONARY BIOLOGY

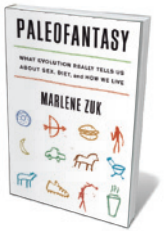
Twisting the tale of human evolution

John Hawks enjoys a debunking of myths about our evolutionary fitness for the twenty-first century.

Advances in medicine and psychology may be stunning, but why is humanity plagued with persistent and widespread ills, from diabetes to depression? Some anthropologists and psychologists speculate that an evolutionary mismatch is at work: we are cave-dwellers struggling in a high-tech world. But as Marlene Zuk explains in *Paleofantasy*, many of these ideas amount to little more than pseudoscience based on an imagined past.

The palaeofantasist's basic idea is this: natural selection tends to make a population well adapted to its environment. But change the environment enough, and finely tuned adaptations start to fail. Human environments in particular have changed radically in the past 10,000 years, meaning that many human traits may not work well today.

Except, argues Zuk, it is not that simple. Zuk, a biologist, reviews how our assumptions about the past have shaped the science of human biology in relation to factors ranging from exercise and diet to mating and marriage. She ably presents a sceptical and light-hearted view of a long list of palaeofantasies and supposed solutions. (My name appears in the book a few times as a sceptic of various poorly supported hypotheses and as a researcher investigating well supported ones.) For instance, our feet seem well adapted for running, yet stress injuries to feet, legs and backs are rife among runners. Palaeofantasy solution: run barefoot. Some psychologists have argued that humans aren't meant for monogamy, because our ancestors were supposed to have interacted like the 'sexually social' bonobo. Palaeofantasy solution: open marriages.



Paleofantasy: What Evolution Really Tells us About Sex, Diet, and How We Live
MARLENE ZUK
W. W. Norton: 2012. 336 pp. £17.99

How do these controversies arise? Because scientists try to reconstruct ancient environments by piecing together archaeological data, comparative primate behaviour and observations on living, small-scale societies such as the Hadza of Tanzania or the Aché of Paraguay that have lifestyles similar to those of our ancestors. But every living and archaeological group has its own distinctive circumstances and history. We can't blithely assume that we can reconstruct the environment relevant to natural selection in the past.

By presenting the state of evolutionary science, Zuk shows that palaeofantasies cannot be justified across a range of environments or with a range of behaviours. She details the evolutionary analysis of human mating patterns, showing that monogamous mating goes way back. Human childhood is long compared with that of our ape relatives, and the palaeofantasy explanation is that cognitive development necessitates a long childhood. Zuk runs through extensive data on the supply side, establishing the credibility of the alternative hypothesis that humans have maintained high rates of reproduction by reducing maternal energy investment in children, instead recruiting grandparents and other relatives to help care for them.

There are other such examples. Many clear cases of recent adaptation show that natural selection has kept pace with some rapid environmental shifts. For example, malaria has induced dozens of genetic adaptations in tropical peoples during the past few thousand years. And diets in the late Pleistocene epoch, which came to a close about 11,700 years ago, were very diverse; Neanderthals, for instance, noshed on cooked grains. With advances in our understanding of the global human population and the rapidly shifting climates of the most recent glaciation, no single model can encompass this diversity.

There are a few real mismatches among the fake ones Zuk highlights. Diet is one. Pleistocene people did not rely on large stored harvests of starchy grains, fatty meat and milk from domesticated animals, or processed sugars — all of which are among the causes of the obesity, type 2 diabetes and metabolic syndrome that are so prevalent today. Some popular diet regimes attempt to rectify the problem by prescribing a menu modelled on one that probably prevailed before the invention of agriculture 10,000 years ago. But although cutting out starchy grains, sugars and milk, and relying on lean meat has helped many people shift to a healthier lifestyle, it has driven some adherents to extremes.

As an anthropologist, I observe that Zuk's use of the term 'fantasy' is just an emphatic way of describing the hypothesis-forming that is essential to evolutionary science. We play with hypotheses, explore their predictions and try very hard to falsify them. So it is, in a way, unremarkable that so many hypotheses proposed by anthropologists about ancient environments now seem to be wrong — and, in a few cases, even ridiculous.

It means that science is working. Genomics, high-resolution climate records, and microscopic and isotopic evidence have changed our understanding of what the past has to offer. With that in mind, let the next round of palaeofantasies begin. ■

John Hawks *is associate professor of anthropology at the University of Wisconsin-Madison, USA, and runs the weblog http://johnhawks.net.*
e-mail: jhawks@wisc.edu

© 2013 Macmillan Publishers Limited. All rights reserved

ARCHAEOLOGY

A distant mirror

Ewen Callaway finds a showing of prehistoric artefacts aesthetically stunning, but a missed scientific opportunity.

"But is it art?" I found the cliché running through my mind as I toured *Ice Age Art*. The aesthetic impact of the scores of pendants, statuettes and other prehistoric objects now showing at London's British Museum is profound. But the show misses an opportunity to burnish the visual impact of the works with an explanation of their importance to human history.

In Europe some 40,000 years ago, *Homo sapiens* — who had arrived from Africa several thousand years earlier — started making objects without any obvious practical use and painting on cave walls. Archaeologists see this as the emergence of 'modern behaviour' and among the first evidence that humans had the capacity for the abstract thinking that underlies language, religion and, yes, art.

The British Museum exhibition argues that even if we will never know why our ancestors created symbolic objects, our shared cognitive capacities mean that we can understand their desire to create art and appreciate the beauty of their creations. It is a good point, and a connection that visitors should appreciate. But in its emphasis on prehistoric objects as fine art and their makers as artists, *Ice Age Art* misses opportunities to explain their historical context and scientific importance.

Art or not, these exhibits are stunning. The collection must represent one of the most comprehensive of prehistoric artefacts ever assembled. The show starts with a bang: a replica of the 40,000-year-old mammoth ivory carving dubbed the Lion Man: a 30-centimetre-high statuette of a man with the head of a lion, discovered in the Stadel cave in southwest Germany in the 1930s. Beside it is a series of delicate miniature ivory animals — a horse, mammoth, cave lion and bison — discovered in the nearby Vogelherd cave and dated to at least 32,000 years ago.

These figures represent some of the earliest known depictions of animals, possibly representing "reincarnated ancestors" or "spiritual helpers", the exhibition notes. Experiments show that the maker may have spent hundreds of hours fashioning the Lion Man, underscoring its importance. A flute of similar age, carved from a vulture bone (and first reported in this publication in 2009), may have emitted "a pleasing sound" if a modern replica is anything to go by. Such playful speculations are always couched as such, and they are based on the interpretations and reconstructions of experts.

➲ NATURE.COM
For a Q&A with Werner Herzog, see:
go.nature.com/3yhkqu



DANIEL PONSARD/MNHN-MH

The Venus of Lespugue, carved more than 20,000 years ago.

Ice Age Art
BRITISH MUSEUM, LONDON
Until 26 May 2013

Left unsaid is that these objects are all linked to the Aurignacian, a stone-tool culture that appears throughout Palaeolithic Europe between about 40,000 and 30,000 years ago. Recent carbon datings support the theory that this culture and its symbolic practices may have emerged in the Swabian Jura region, where the objects were discovered, before fanning out through Europe.

Together, the objects indicate a population undergoing rapid cultural upheaval compared with their ancestors. But there is no mention that scientists propose population density and climate change as common catalysts for such cultural explosions. Such context wouldn't detract from the objects' beauty, and it would give visitors a glimpse of why such discoveries are so important to our understanding of this period. And the exhibition barely mentions that humans living in southern Africa more than 80,000 years ago also exhibited symbolic behaviour, as found in the Blombos cave (although the objects are fewer and very much simpler).

The aesthetic link between ancient symbolic objects and modern art is made explicit by including alongside them works by abstract artists such as Henri Matisse, Henry Moore and Pablo Picasso. These strike me as a distraction. Indeed, both Matisse and prehistoric humans abstracted the female form into simple lines and shapes — a prehistoric wishbone-shaped ivory pendant from Dolní Věstonice in the Czech Republic is one of my favourites — but so does the sign outside the ladies' toilet.

The exhibition's nod to more modern art is most effective when used metaphorically. For example, the period after the most recent glacial maximum around 20,000 years ago is called a 'Renaissance' because humans began exploring new forms of figurative expression, which include stone animal engravings that play with shading and perspective, and the cave paintings of Lascaux in France.

Cave art is difficult to present in any museum. But even so, the context-less montage of images from Lascaux and other prominent caves in a dark room at the exhibition feels more like a modern-art installation. Instead, watch Werner Herzog's spellbinding documentary *Cave of Forgotten Dreams*, filmed in the Chauvet-Pont D'Arc cave in France. Better still, go to France and experience the real thing.

The press release that accompanies *Ice Age Art* states that the objects are presented as art and not as archaeological finds. It is a shame, and a puzzle, that they can't be both. ■

Ewen Callaway *is a reporter for* Nature.

© 2013 Macmillan Publishers Limited. All rights reserved

Correspondence

Integrity rehab leader responds

As leader of the programme for Restoring Professionalism and Integrity in Research (RePAIR; see *Nature* **493,** 147; 2013), I wish to correct some popular assumptions about its approach.

The RePAIR programme offers remediation for researchers who have been persistently non-compliant or engaged in wrong-doing. It is not an alternative to preventing non-compliance in the first place. Neither is it suitable for all violators: before referring an investigator, an institution must believe that he or she is talented and capable of reform.

The programme was not designed primarily for researchers guilty of falsification, fabrication or plagiarism. All those referred so far have failed to protect human subjects, to care for animals properly, or to declare and manage conflicts of interest.

Most investigators remain in their field after disciplinary action. RePAIR responds to this fact, partly to avoid compromising the future of co-workers (see J. M. DuBois *et al. Clin. Transl. Sci.* http://doi.org/kp6; 2013).

Complex compliance demands make researchers more prone to falling foul of the rules, particularly when they are under stress from time constraints and funding pressures. More education in ethics is not the answer: instead, proper strategies are needed for managing stress, solving professional problems and dealing with obstacles to compliance. For this reason, the RePAIR programme was designed not by ethicists but by a team of specialized psychologists.

James M. DuBois *Saint Louis University, St Louis, Missouri, USA.*
duboisjm@slu.edu

Funding: some sense in overlapping grants

I believe it makes sense in these straitened times to submit scientifically overlapping grant applications to multiple agencies, in the hope that one will receive funding (*Nature* **493,** 577, 2013).

It would, of course, be unethical to accept duplicate budgets for the same salary or equipment. But overlapping funding is often used to expand and upgrade laboratory resources, particularly if one grant provides insufficient support to follow up the scientific leads and questions that always pop up along the way.

A simpler, more efficient system could be achieved with more overlap, not less. One solution might be a centralized proposal repository (N. Y. Harel *Nature* **452,** 409, 2008), into which individual researchers could submit one proposal for review by multiple agencies. Several agencies might then share the costs of (and credit for) the same project.

Instead of sleuthing to identify other agencies that received similar proposals, funders could use text-similarity software to find and collaboratively fund the proposals that are most relevant to their missions. A system of overlapping funding would also encourage more transparency.

Noam Y. Harel *Icahn School of Medicine at Mount Sinai; and James J. Peters Veterans Affairs Medical Center, New York, USA.*
noam.harel@mssm.edu

Funding: database to counter duplication

An initiative by the International Cancer Research Partnership (ICRP; chaired by S.B.R.) is helping to address the problem of duplicate funding for similar research (*Nature* **493,** 577; 2013 and *Nature* **493,** 599–601; 2013).

The ICRP represents 56 cancer-research organizations in the United States, Canada, the United Kingdom, France, the Netherlands, Australia and Japan. It curates a publicly accessible database of details of each partner's funding and research (see go.nature.com/ckanbr).

Around one-third of our partners ask grant applicants to interrogate this database for overlapping and complementary research. Other partners use the database as an internal administrative tool to assess potential duplication of effort. The process is straightforward because all awards are structured using an international classification system known as the Common Scientific Outline.

At the ICRP, we believe that cancer research should be a coordinated and collaborative global enterprise, with maximally efficient use of scarce resources. We therefore invite other cancer-research organizations to join us in order to meet these objectives.

Stephanie Birkey Reffey *Susan G. Komen for the Cure, Washington DC, USA.*
Lynne Davies *ICRP, Cancer Research UK, London.*
lynne.davies@cancer.org.uk

Marine monitoring is hard and costly

While fisheries scientists are still struggling with uncertainties in stock assessments for single species (*Nature* **494,** 282; 2013), a giant leap towards a whole-ecosystem approach is being embedded into international policies for fisheries management. However, underwater monitoring of a huge and complex range of environmental and biological processes will be difficult and costly.

The unsustainable use of a resource can have knock-on effects within an ecosystem, as well as socioeconomic repercussions. If budget constraints limit the monitoring of whole communities, however, population estimates and predictions could be compromised, possibly masking a species decline and leading to misguided policy decisions. Governments should not be tempted to find inexpensive solutions to fulfil policy requirements and to meet deadlines.

There are tools available to help countries pinpoint problems, direct research efforts, quantify acceptable limits for degradation, calculate the monitoring effort needed to detect a meaningful change, and gather resources for implementing a useful strategy (see B. S. Halpern *et al. Nature* **488,** 615–620; 2012).

Miguel Pessanha Pais *Centre for Oceanography, University of Lisbon, Portugal.*
mppais@fc.ul.pt

Use foreign aid to help African science

Scientific and technological research could solve many of Africa's local issues, but these problems persist despite generous international support.

The situation is made worse by a severe shortage of well-trained, home-grown scientists, and aggravated by emigration and poor working environments (see, for example, M. H. Hassan *Cell* **131,** 433–436; 2007). Countries in Africa lack adequate equipment, have no running budgets and almost no ethics codes or regulations. These factors also discourage African scientists trained abroad from returning home and applying their valuable skills.

Solutions derived from the developed world are unlikely to work, because Africa's reality is centuries behind and there is no infrastructure for implementing them. It is a matter of priority to bridge this yawning gap, otherwise the continent could be left behind for ever.

Africa should be using some of its international aid to guarantee long-term education in science and technology for new generations: only then can our research be self-sufficient and our science capacity advance.

Khalid D. Awadelkarim *National Cancer Institute, University of Gezira, Wad Medani, Sudan.*
awadelkarim@gmail.com

© 2013 Macmillan Publishers Limited. All rights reserved

PUBLIC HEALTH

Broad reception for coronavirus

The discovery that a new coronavirus associated with lethal respiratory infections binds to an evolutionarily conserved receptor on airway cells suggests that direct transmission from bats to humans may occur. SEE LETTER P.251

TOM GALLAGHER & STANLEY PERLMAN

Among the vast pool of zoonotic viruses — those that can be transmitted from animals to humans — are the coronaviruses. These pathogens cause common colds and, rarely, more severe infections, and their entrances into the human population provide case studies in virus evolution and the emergence of infectious disease. A new coronavirus, designated hCoV-EMC, has appeared in humans in Middle Eastern countries and England on several occasions over the past year. Although only a few people have been infected, around half of them have died[1]. The identification of the receptor for this virus, reported by Raj *et al.*[2] on page 251 of this issue, will help us to understand how this and other coronaviruses can cross species borders.

Coronaviruses made the headlines during the 2002–03 epidemic of human severe acute respiratory syndrome (SARS). The origin of the pathogenic human virus (SARS-CoV) that caused these infections was traced back to infected bats[3]. In expanding its range to humans, the SARS-like bat viruses evolved the capacity to latch onto human lung-cell receptors[4]. This probably came about through infections in intermediate host animals, in which genetic recombinations and small mutations generated a human-adapted virus that could no longer infect bat cells. The infrequency and requisite specificity of such events may partly explain why we have not seen another SARS-like coronavirus epidemic in the past decade.

The emergence of hCoV-EMC, which is associated with severe respiratory disease, multi-organ failure and a high mortality rate in patients, has prompted intensive investigation into the virus's replication and cell-targeting strategies. Once again, its closest known relatives are viruses that infect various bat species[5]. However, in stark contrast to SARS-CoV, hCoV-EMC indiscriminately infects cells from many sources, including bats, pigs, monkeys and humans[6]. A polytropic coronavirus of this type is highly unusual and, from an epidemiological perspective, seems alarming. The suggestion is that hCoV-EMC has acquired facile interspecies transmissibility by adapting to evolutionarily conserved host-cell components — including host-cell receptors.

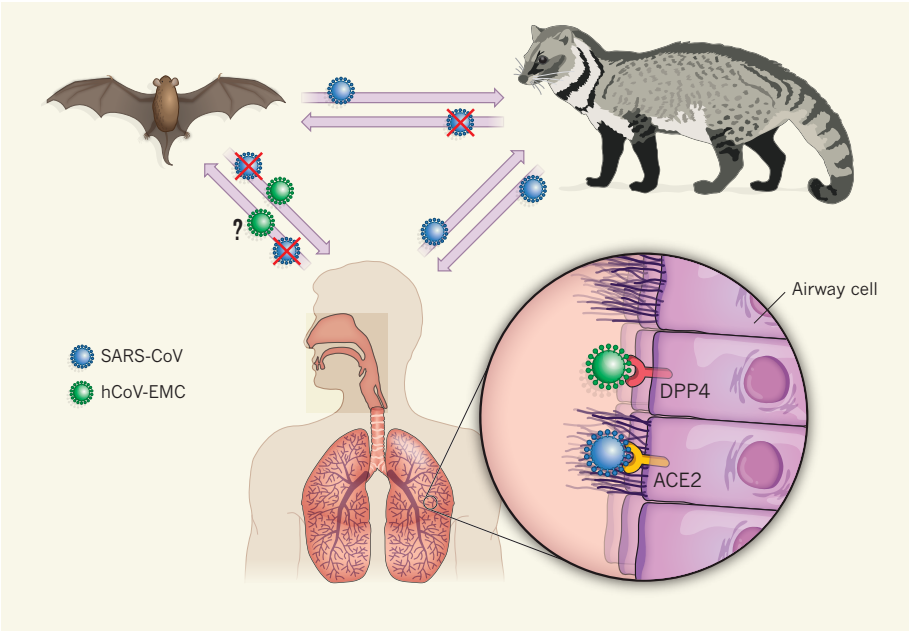


Figure 1 | Interspecies transmission pathways for coronaviruses. Raj and colleagues[2] have identified the cell-surface protein DPP4 as the receptor for hCoV-EMC, a new coronavirus that causes severe respiratory infections in humans. DPP4 is expressed on non-ciliated cells in the human airway. The virus is also able to use the homologous protein in bats for infection, which suggests that direct and reversible transmission of the virus between bats and humans may occur (although transmission through an intermediate host remains a possibility). By contrast, another pathogenic coronavirus (SARS-CoV), which binds to the ACE2 receptor on ciliated airway cells, probably cannot be transmitted directly and is likely to have jumped from bats to humans through evolutionary processes in intermediate hosts, such as civet cats. Open arrows indicate putative routes of transmission, crosses indicate evidence against transmission and a question mark indicates speculation on transmission.

Raj and colleagues' results validate this idea. They found that hCoV-EMC binds to DPP4 (also called CD26), a protein found on the surface of several cell types, including cells in the human airways. The amino-acid sequence of human DPP4 is very similar to its homologues in *Pipistrellus pipistrellus* bats, and the authors demonstrate that bat DPP4 also functions as a hCoV-EMC receptor (Fig. 1). The sequence similarities extend to the DPP4s of other wild and domesticated animals, which increases the likelihood that dissemination of hCoV-EMC throughout nature involves the use of DPP4 for cell entry.

Insight into whether hCoV-EMC has adapted to bind exclusively to an interspecies-conserved domain of DPP4 must await characterization of the binding interface, which will require mutational analyses of DPP4 and the glycoproteins on the hCoV-EMC surface. These analyses will be guided in part by comparisons with the closest known hCoV-EMC relatives — bat coronaviruses HKU4 and HKU5. In addition, X-ray crystallography of receptor-binding domains of viral proteins in complex with soluble DPP4 will be especially enlightening. Dissecting the hCoV-EMC receptor-binding domain and its interacting DPP4 domain in these ways may reveal opportunities for therapeutic blockade of virus entry into cells, including the development of vaccines based on the receptor-binding domains that can elicit antibodies able to interfere sterically with viral entry.

Beyond the potential therapeutic applications, the identification of the hCoV-EMC receptor may also provide clues about the virus's pathogenesis. This is because the DPP4

© 2013 Macmillan Publishers Limited. All rights reserved

receptor resembles the α-coronavirus receptor APN and the SARS-CoV receptor ACE2. All three receptors are ectopeptidase enzymes that cleave amino acids from biologically active peptides, thereby regulating an array of physiological responses. However, APN, ACE2 and DPP4 do not share obvious structural features, and their peptidase activities are not necessary for virus entry[7,8]. Coronavirus adaptation to ectopeptidase receptors may, therefore, simply reflect the abundance or subcellular positioning of these enzymes on airway cells.

That said, once they have robustly infected a cell, viruses interfere with the presentation of such receptors at the cell surface, and decreased ACE2 levels during SARS-CoV infection are correlated with enhanced disease severity[9]. Further research to determine whether hCoV-EMC disease is similarly related to dysregulated DPP4-mediated physiological responses will address the intriguing hypothesis that aspects of coronavirus pathogenesis are outcomes of adaptation to ectopeptidase receptors.

Knowing the identity of the hCoV-EMC receptor will also allow the development of animal models of infection to assess whether there are causal links between DPP4 levels, hCoV-EMC infection and disease. For example, an evaluation of DPP4 distribution in the lungs will help to show whether the receptors' location restricts hCoV-EMC infections to the lower respiratory tract, which might limit the virus's transmissibility. DPP4 is found on non-ciliated airway cells, whereas ACE2 is expressed by ciliated cells (Fig. 1); such cell-target differences may contribute to the differences in transmissibility and pathogenicity of infections caused by hCoV-EMC and SARS-CoV. The potential for other factors — such as soluble DPP4, which may be abundant in extracellular fluids — to preclude infection and disease should also be tested. Moreover, DPP4 is known to have roles in recruiting protective immune responses in the host[10]; as such, the effects of virus-induced receptor dysregulation may feature prominently in elucidating immunopathological aspects of the disease.

Although hCoV-EMC can be transmitted from human to human, fortunately this seems to occur infrequently. Further epidemiological studies should assess whether the human infection is truly rare and always severe or, alternatively, is widespread but generally mild and therefore not detected. Similar epidemiological considerations apply to animals. Although the immediate implication of Raj and colleagues' findings might be to postulate a direct transmission from bats to humans, the conservation of the DPP4 receptor among species also raises questions about the extent of hCoV-EMC in nature and the most proximal animal source of the human infections. The interesting and perhaps troubling findings from studies of this virus thus far are that there may be a plethora of sources for its intrusion into the human population. Is this the case, or are there distinct interspecies barriers to hCoV-EMC infection? If so, what is the nature of the barriers, and how might the virus adapt to cross them and occupy the human-host niche? Virus adaptations involve much more than evolving new receptor-binding domains and so, in further studies of this emergent pathogen, it will be important to consider other genetic determinants of hCoV-EMC transmission to humans, such as virus interactions with the innate immune system. ■

Tom Gallagher *is in the Department of Microbiology and Immunology, Loyola University Medical Center, Maywood, Illinois 60153, USA.* **Stanley Perlman** *is in the Department of Microbiology, University of Iowa, Iowa City, Iowa 52242, USA.*
e-mails: tgallag@lumc.edu; stanley-perlman@uiowa.edu

1. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. *N. Engl. J. Med.* **367,** 1814–1820 (2012).
2. Raj, V. S. *et al. Nature* **495,** 251–254 (2013).
3. Li, W. *et al. Science* **310,** 676–679 (2005).
4. Becker, M. M. *et al. Proc. Natl Acad. Sci. USA* **105,** 19944–19949 (2008).
5. van Boheemen, S. *et al. mBio* **3,** e00473-12 (2012).
6. Kindler, E. *et al. mBio* **4**, e00611-12 (2013).
7. Reguera, J. *et al. PLoS Pathogens* **8,** e1002859 (2012).
8. Li, F., Li, W., Farzan, M. & Harrison, S. C. *Science* **309,** 1864–1868 (2005).
9. Kuba, K. *et al. Nature Med.* **11,** 875–879 (2005).
10. Ansorge, S. *et al. Clin. Chem. Lab. Med.* **47,** 253–261 (2009).

EARTH SCIENCE

Core composition revealed

The composition of Earth's core may be easier to resolve than previously thought. Laboratory experiments strengthen the hypothesis that oxygen and silicon are the prime candidates for the light elements present in the outer core.

LIDUNKA VOČADLO

How do you work out the composition of Earth's core when it lies 3,000–6,000 kilometres below your feet and is completely inaccessible? Writing in *Geophysical Research Letters*, Tsuno and colleagues[1] have tried to do just that by using a series of laboratory experiments to recreate the conditions expected during the formation of the core.

From studies involving seismology and cosmochemistry, we know that Earth's core consists predominantly of iron, alloyed with up to about 10% nickel. But the density of the core as measured from seismic-wave velocities is significantly lower (by 10–15%) than that of a pure metal alloy, and therefore the core must also contain some light elements. In a dynamic process, the solid inner core is crystallizing out of the liquid outer core, preferentially expelling light elements into the outer core. The inner core contains about 3% light elements, whereas the outer core contains nearer 10%. Cosmochemical and other arguments suggest that obvious candidates for these light elements include sulphur, silicon, oxygen and carbon, and these elements (and others) have been studied extensively for the past 20 years or so. But none of these light elements seems to satisfy all the requirements imposed by seismology, cosmochemistry and the evolutionary models describing how Earth has evolved over time.

Given that the liquid outer core is surrounded by a silicate lower mantle, it is not surprising that silicon and/or oxygen are prime candidates for the light elements in the outer core. Also, a few per cent silicon hidden in the core conveniently accounts for the observation that silicon is slightly lower in the silicate part of Earth than in chondritic meteorites, which are thought to reflect the primordial composition of the whole Earth.

Further support for silicon and/or oxygen comes from arguments based on volatility data that tend to reject the more volatile carbon and sulphur (among others) in favour of silicon. In the case of oxygen, however, there has been a problem, because geochemical models of Earth's early evolution suggested that the core could only have formed with little or no oxygen[2]. However, a recent study[3] indicates that, conversely, the core could actually have formed under oxidizing conditions, thereby making oxygen a viable candidate for a light element in the core after all.

But having only silicon or only oxygen in the outer core is not sufficient to explain the core's composition. A theoretical study[4] has shown that, at the pressure and temperature conditions of the present-day boundary between Earth's inner and outer cores (a pressure of 330 billion pascals and an estimated

© 2013 Macmillan Publishers Limited. All rights reserved

temperature of up to 6,000 kelvin), silicon is equally able to be present in the solid inner core and the liquid outer core. Thus, silicon will not produce the marked change in density that is observed from seismology studies of the boundary between the solid and liquid core. By contrast, oxygen can exist only in the liquid outer core, and thus will produce a density jump at this boundary. Therefore, to obtain the required reductions in density for both the inner and outer cores, together with the density jump across the boundary between them, both of these elements seem to be necessary.

Nonetheless, having oxygen and silicon together in the core poses another problem. When experimentalists try to put oxygen and silicon into liquid iron, their results indicate that these elements seem to be mutually exclusive — in fact, in the steel industry, silicon has been used as a deoxygenating agent. Furthermore, this incompatibility has been found to persist at the high pressures (40–60 GPa) and temperatures (less that about 3,000 K) expected during core formation in the early stages of Earth's evolution.

Tsuno and colleagues' work may have provided a solution to this problem of core composition. The authors performed experiments at a pressure of 25 GPa and a temperature of about 3,000 K on samples consisting of a liquid iron-rich metal interfaced with liquid silicate. The samples also included iron oxide, as this is particularly important in controlling the way in which silicon and oxygen partition between the silicate and the metal. They then determined the distribution of silicon and oxygen between the silicate and the metal alloy, and used the results to develop a thermodynamic model of the system. This model predicts that, when the temperature is increased by only a few hundred degrees (from 3,000 K to 3,500 K), the amounts of silicon and oxygen present in liquid iron increase to well above those required to match the density deficit inferred from seismic-wave data. Although the conclusions from this work have been suggested before[5], it is the incorporation of iron oxide, which controls the availability of oxygen in the experiments, that makes the authors' conclusions more definitive.

However, it is worth noting that Tsuno and colleagues' experiments were performed at relatively modest temperatures (about 3,000 K), and so their conclusion that silicon is more readily incorporated into liquid iron alloyed with oxygen at 3,500 K is based on an extrapolation of their data. This will need to be tested. Moreover, a requirement for oxygen and silicon to be the light elements in the core is that this iron–oxygen–silicon alloy is then able to match the observed seismic-wave velocities and the density of the present-day core. The relevant physical properties of the alloy, however, are not yet known at the temperatures and pressures of the present-day core.

Finally, the samples used here contained no nickel and, at the temperatures of these experiments, nickel may not be as iron-like (that is, have similar physical and chemical properties) as previously thought[6]. Nevertheless, the arguments for a combination of oxygen and silicon as being the light elements in the core have now been significantly strengthened. Perhaps the most likely and most abundant elements do provide the real solution after all. ■

Lidunka Vočadlo *is in the Department of Earth Sciences, University College London, London WC1E 6BT, UK.*
e-mail: l.vocadlo@ucl.ac.uk

1. Tsuno, K., Frost, D. J. & Rubie, D. C. *Geophy. Res. Lett.* **40,** 66–71 (2013).
2. Wood, B. J., Walter, M. J. & Wade, J. *Nature* **441,** 825–833 (2006).
3. Siebert, J., Badro, J., Antonangeli, D. & Ryerson, F. J. *Science* http://dx.doi.org/10.1126/science.1227923 (2013).
4. Alfè, D., Gillan, M. J. & Price, G. D. *Earth Planet. Sci. Lett.* **195,** 91–98 (2002).
5. Takafuji, N., Hirose, K., Mitome, M. & Bando, Y. *Geophys. Res. Lett.* **32,** L06313 (2005).
6. Martorell, B., Brodholt, J., Wood, I. G. & Vočadlo, L. *Earth Planet. Sci. Lett.* **365,** 143–151 (2013).

CELL BIOLOGY

Alternative energy for neuronal motors

Neurons use molecular motors to power the transport of cargoes along their axonal extensions. Fresh evidence challenges the view that cellular organelles called mitochondria are the main energy providers for this process.

GIAMPIETRO SCHIAVO & MIKE FAINZILBER

Intracellular transport requires energy, and nowhere is this process more energy-demanding than in the axon of a neuron. The length of an axon can range from millimetres in flies to metres in large mammals — a prodigious distance for the molecular motors that travel back and forth conveying cellular cargoes between the neuronal cell body and the nerve endings. These motors move along microtubules in discrete steps of a few nanometres each, and every step requires energy that the motors derive from the hydrolysis of ATP molecules. Consequently, axonal transport consumes large amounts of ATP, and organelles called mitochondria are widely assumed to be the principal source of this energy. But in a paper published in *Cell*, Zala *et al.*[1] challenge this assumption, showing that a sugar-hydrolysing enzyme bound to the surface of transport vesicles provides the energy for this process. In other words, motor complexes in axons seem to use on-board 'generators' to ensure a steady energy supply en route to their destination.

Mitochondria regulate their mobility according to the intracellular levels of calcium ions that accumulate in the axon at specialized energy-requiring sites rich in ion channels[2]. The distance between adjacent clusters of mitochondria can be greater than known ranges of mitochondrial-ATP gradients[3]. Zala *et al.* therefore set out to identify the energy source for transport in axonal regions that lack mitochondria. Surprisingly, inhibition of mitochondrial ATP production did not significantly affect the axonal transport of vesicles carrying the growth factor BDNF, regardless of the apparent ATP levels in the axon.

Cells can generate energy independently of mitochondria by breaking down sugars using a process called glycolysis. Unexpectedly, Zala and colleagues found that chemical inhibition of the key glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH), or reduction in its expression, blocked the transport of axonal vesicles. GAPDH was present on the surface of these transport vesicles (Fig. 1), where it was bound by huntingtin — a motor-associated protein that is mutated in the neurodegenerative disorder Huntington's disease[4]. These findings have strong parallels with earlier work[5] demonstrating a role for GAPDH, but not mitochondria, in the bioenergetics of the uptake of neurotransmitter molecules into synaptic vesicles at nerve endings. The two studies add localized energy provision to the growing list of diverse functions of GAPDH in different subcellular compartments[6].

Zala and co-authors' work might also have therapeutic implications. Their results indicate that mitochondrial impairment is unlikely to be the main cause of the alterations in fast axonal transport seen in neurodegenerative disorders such as Huntington's disease and amyotrophic lateral sclerosis (ALS). This could explain the failure of previous efforts[7] to ameliorate disease in a mouse model of ALS by increasing mitochondrial mobility. Furthermore, the affinity of GAPDH for aggregating proteins found in various neurodegenerative disorders[8] indicates that sequestration of this enzyme away from transport vesicles

© 2013 Macmillan Publishers Limited. All rights reserved

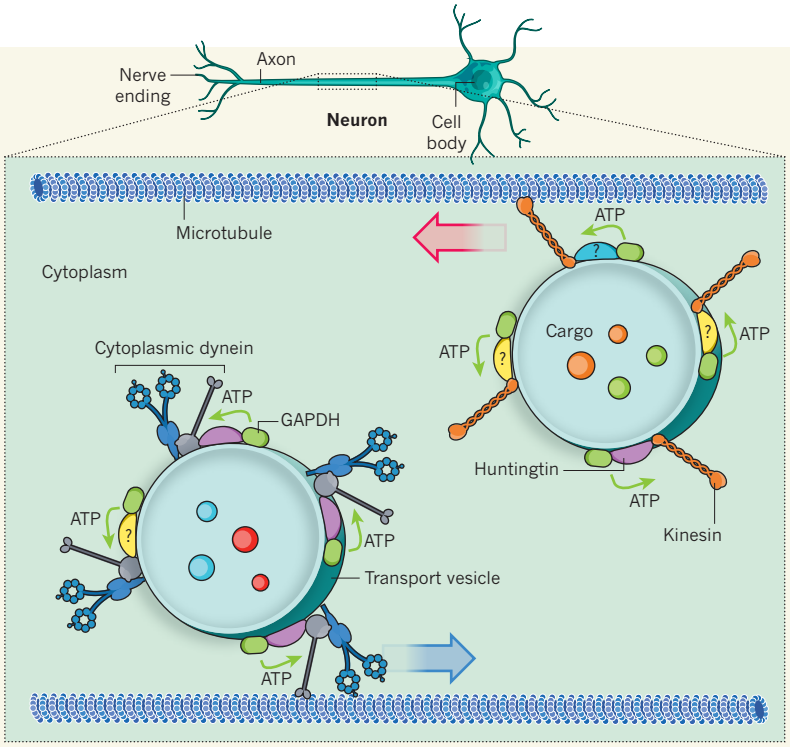


Figure 1 | On-board energy generators. Neurons have long axonal processes that require a constant supply of energy in the form of ATP molecules to transport cargo-carrying vesicles along microtubules from the cell body to axonal terminals (red arrow) and back (blue arrow). The prevalent assumption is that mitochondria (not shown) are the source of ATP for this purpose. Zala *et al.*[1], however, find that the glycolytic enzyme GAPDH, which is bound to transport vesicles through the huntingtin protein and possibly other factors (question marks), powers the motor proteins cytoplasmic dynein and kinesin on microtubules by locally generating ATP through glycolysis.

might contribute to disease development and progression. Nonetheless, therapeutic targeting of GAPDH will probably require the development of drugs that specifically modulate only a subset of its functions[6], as indicated by the failure of a GAPDH-directed molecule in a clinical trial for treating ALS[9].

The uncoupling of vesicular axonal transport from mitochondrial ATP production is the most intriguing yet puzzling aspect of this work. For instance, Zala *et al.* describe an alternative energy source for fast transport across axonal areas devoid of mitochondria, yet they did not detect any overt variation in ATP levels in the axons they analysed. Another mystery arises from the authors' observations that mitochondrial ATP plays no part in fast axonal transport, and that motors associated with vesicular cargoes cannot access this source of ATP. This finding cannot be explained by intrinsic properties of the motor complexes, because the same types of molecular motor can hydrolyse mitochondrial ATP for mitochondrial transport, which was not affected by inhibition or reduced expression of GAPDH[1].

Similarly, a direct link between GAPDH and motor complexes is unlikely. The authors report that GAPDH association with transport vesicles is necessary for axonal transport, although its binding to molecular motors is not: in huntingtin-depleted cells, GAPDH targeting to vesicles by other means was sufficient to recover transport. A direct link between GAPDH and motor complexes is also unlikely because GAPDH-mediated axonal transport was shown to occur with different motors moving in different directions (Fig. 1). To resolve this conundrum, *in vitro* studies are required that examine the effects of GAPDH on the biophysical properties of isolated motor complexes moving along microtubules. Such studies might reveal additional functions for GAPDH in regulating these molecular nanomachines. For example, it could be that other enzymatic activities of GAPDH, such as *S*-nitrosylation and mono-ADP-ribosylation[6,8], are also necessary for transport.

Whether GAPDH is required for the axonal transport of non-vesicular complexes, such as RNA granules, is also unknown. And does a similar GAPDH-dependent mechanism apply to other cellular ATP-hydrolysing enzymes, such as ion pumps and remodelling enzymes[8]? Whatever the answers, Zala and co-workers' findings place alternative energy sources in the axonal domain, calling for a re-examination of fundamental assumptions and pointing the field in intriguing and unexplored directions. ■

Giampietro Schiavo *is at the Cancer Research UK London Research Institute, London WC2A 3LY, UK.* **Mike Fainzilber** *is in the Department of Biological Chemistry, Weizmann Institute of Science, 76100 Rehovot, Israel.*



50 Years Ago

Mr. A. J. Kirkman has recently argued in *Nature* that it was mainly up to teachers of science and technology to give the guidance and impose the standards necessary for their students to improve their written English … All the errors he mentioned were fairly easy to spot, easy to correct, and of a kind which the teacher could easily mark; but they were all irrelevant to the main problem of helping scientists to write better English and express themselves more fluently. The damage is done in the classes called 'English', and is not the fault of science teachers who accept low standards … One is led to suspect that a teacher of English can hope to find only the superficial deficiencies in scientific English unless he learns more about the subject-matter. Perhaps Mr Kirkman is right after all in passing responsibility to the science teachers because most English teachers cannot bear it. But the scientist's task is so often rendered well-nigh impossible by English teachers who inoculate pupils interested in science with a distaste, or at least a disdain, for the subject known as 'English'.

From *Nature* 16 March 1963

100 Years Ago

Some readers of *Nature* will be interested to learn that tadpoles with large suctorial oral discs, enabling their possessors to adhere firmly to the rocks and boulders of mountain streams, have recently been discovered at Krantzkloop, in Natal, at an elevation of about 1500 to 1600 ft. They were found by the Rev. Fr. P. Boneberg, of Mariannhill, who kept them alive for some time, and observed their peculiar leech-like habit of sticking to one's fingers or to the sides of the vessel in which they were contained.

From *Nature* 13 March 1913

© 2013 Macmillan Publishers Limited. All rights reserved

e-mails: giampietro.schiavo@cancer.org.uk; mike.fainzilber@weizmann.ac.il

1. Zala, D. *et al. Cell* **152,** 479–491 (2013).
2. MacAskill, A. F. & Kittler, J. T. *Trends Cell Biol.* **20,** 102–112 (2010).
3. Niethammer, P., Kueh, H. Y. & Mitchison, T. J. *Curr. Biol.* **18,** 586–591 (2008).
4. Caviston, J. P. & Holzbaur, E. L. *Trends Cell Biol.* **19,** 147–155 (2009).
5. Ikemoto, A., Bole, D. G. & Ueda, T. *J. Biol. Chem.* **278,** 5929–5940 (2003).
6. Tristan, C., Shahani, N., Sedlak, T. W. & Sawa, A. *Cell Signal.* **23,** 317–323 (2011).
7. Zhu, Y.-B. & Sheng, Z.-H. *J. Biol. Chem.* **286,** 23432–23440 (2011).
8. Seidler, N. W. *Adv. Exp. Med. Biol.* **985,** 249–267 (2013).
9. Miller, R. *et al. Neurology* **69,** 776–784 (2007).

MATERIALS SCIENCE

Porosity in a single crystal

A low-temperature synthesis has been developed to make single crystals of titanium dioxide that contain pores tens to hundreds of nanometres in size. This opens the way to cheap, highly efficient optoelectronic devices. SEE LETTER P.215

CATERINA DUCATI

The discovery that nanoparticles of metal-oxide semiconductors can be used to make highly efficient solar-cell devices using low-cost processes has attracted a tremendous research effort worldwide[1,2]. One of the main challenges is to make semiconductors that have a high surface area, but maintain good charge transport. Nanoparticles have high surface areas, but when used in the composite materials that are required for third-generation solar cells, a large number of interfaces are generated between particles; the more interfaces there are, the greater is the hindrance of electron transport across the nanoparticle network, and the less efficient is the device. In this week's issue, Crossland *et al.*[3] (page 215) report that this problem can be solved if the network is replaced by a porous single crystal of a semiconductor*.

Ideal mesoporous single crystals (MSCs) have an ordered atomic lattice that develops around cavities of sizes in the range of tens to hundreds of nanometres. Until now, methods for preparing mesoporous materials — particularly semiconductors — have achieved only porous assemblies of nanometre-scale crystals.

*This article and the paper under discussion[3] were published online on 6 March 2013.

REF. 3

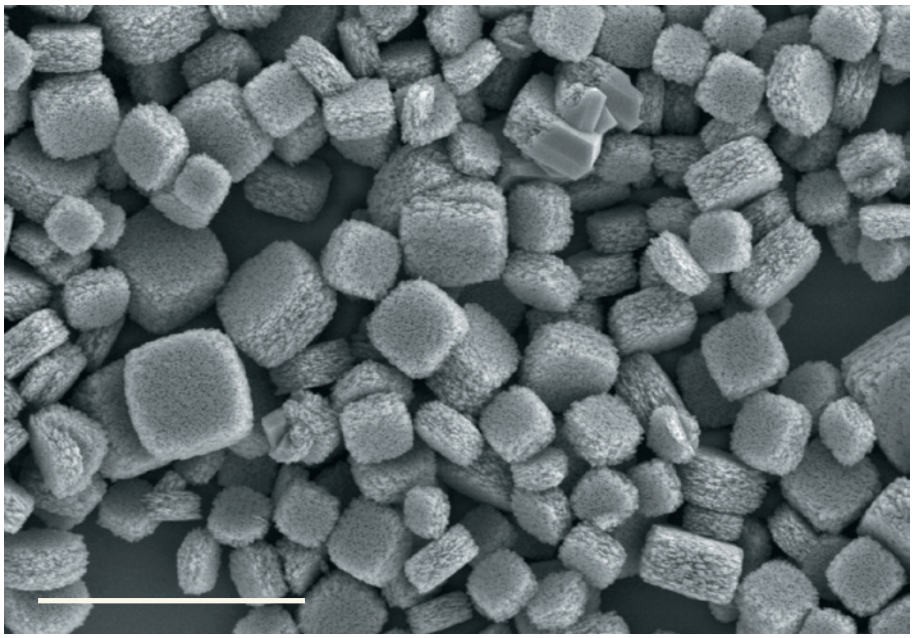


Figure 1 | Mesoporous single crystals. This micrograph shows some of Crossland and colleagues' mesoporous single crystals of titanium dioxide[3], which might open up applications in optoelectronics. Scale bar, 5 micrometres.

By contrast, Crossland and colleagues' technique yields MSCs of titanium dioxide (TiO_2) up to 1 micrometre in size, but with high surface areas because of their large porosity.

This is important in solar-cell applications because the high surface area of the MSC maximizes the probability of generating free electrons, and the large crystal size means that electrons travelling through an MSC-based active layer (the photoanode) of a solar cell would cross just 2–5 interfaces. By comparison, electrons travelling through a photoanode containing 20-nanometre-sized particles might have to cross as many as 50 interfaces. The chances of an electron generating a useful current are therefore much higher in the MSC system.

The key to the success of Crossland and colleagues' method is tight control of nucleation (localized formation of tiny crystals that act as 'seeds' for crystallization processes) and of crystal growth. Crossland *et al.* prepared a 'sacrificial' template structure from close-packed arrays of silica beads, then treated it with titanium tetrachloride to seed it with TiO_2 nuclei[4]. The template was then exposed to a solution of titanium tetrafluoride, using a proven recipe[5] for synthesizing TiO_2 in its 'anatase' crystal form. Finally, they removed the template using a selective etching process, leaving behind porous crystals of TiO_2 (Fig. 1). The networks of pores in the crystals were ordered and interconnected, and the size of the cavities was determined by the size of the beads in the sacrificial template.

Crucially, their approach works at low temperatures: the highest temperatures required are 500 °C, with the main steps requiring 210 °C at most. This should allow the crystals to be integrated into plastic substrates, and minimize the environmental impact and production costs of the material.

The authors also performed a detailed analysis of their synthetic process to study the effects of seed density, temperature and size of the silica spheres on the size and morphology of the resulting crystals (see the Supplementary Information for the paper[3]). Crossland *et al.* then provided the first demonstration of the electronic properties of anatase TiO_2 MSCs, and of the crystals' photovoltaic properties — their ability to convert light into electrical power. Even though these are only preliminary results, the authors show that MSCs do favour electron transport and are characterized by higher electron mobilities than nanoparticle networks.

Internal surfaces of MSCs probably act as electron-scattering centres, thereby complicating electron transport. Further investigations might show how porosity affects such transport. If, as the authors hint, it is possible to explore different template geometries (for example, by making templates from self-assembling polymers), then it may be possible to optimize surface area and

© 2013 Macmillan Publishers Limited. All rights reserved

charge-carrier transport independently.

An exciting development that will probably follow on from Crossland and co-workers' study is the exploration of methods to fine-tune the electronic and optical properties of TiO_2 MSCs. In both cases, it may be beneficial to substitute oxygen atoms in the crystal lattice for other atoms, or to coat the internal surfaces with appropriate (metallic) overlayers. One of the immediate benefits of adding other elements to anatase TiO_2 MSCs is that it may then be possible to modify the crystals' light-absorption properties, allowing the material to respond to visible light rather than just to ultraviolet light. This has obvious implications for photovoltaic and catalytic applications of TiO_2 nanoparticles.

Crossland *et al.* show that their MSCs can be coupled with light-harvesting materials (known as light sensitizers) to make solar cells in which all the components were prepared at temperatures below 150 °C. These cells achieved a record light-to-electricity power conversion efficiency of 7.2%, although this was partly as a result of using a state-of-the-art light sensitizer[6]. The cells' performance is markedly better than that of nanoparticle-based systems that are processed at similarly low temperatures.

The development of low-temperature processes for making optoelectronic devices has already enabled the construction of transparent, flexible and portable gadgets, such as smartphones and tablets, and will facilitate many other important advances. For example, scientists are developing inexpensive solar cells that can be integrated into the fabric of buildings. These devices need to be efficient, light, unobtrusive, self-contained, easily maintained and readily combined with existing architectural elements. MSCs of anatase TiO_2 might be ideal components for these cells.

Although the successful synthesis of TiO_2 MSCs allows us to speculate on how they might be used in energy-harvesting applications, fundamental questions remain to be answered. How are the internal concave surfaces of the crystals structured? What material defects ensue from the unusual geometry of the crystals? And could this geometry affect the chemistry of TiO_2 — especially its chemical response to light — in ways that might improve the material's performance for other applications, such as hydrogen production and the degradation of toxic substances and pollutants in the environment? If Crossland and colleagues' synthetic method can be scaled up to meet potential demands, then the significance and impact of TiO_2 MSCs could be tremendous. ■

Caterina Ducati *is in the Department of Materials Science and Metallurgy, University of Cambridge, Cambridge CB2 3QZ, UK.*
e-mail: cd251@cam.ac.uk

1. O'Regan, B. & Grätzel, M. *Nature* **353,** 737–740 (1991).
2. Grätzel, M. *Acc. Chem. Res.* **42,** 1788–1798 (2009).
3. Crossland, E. J. W. *et al. Nature* **495,** 215–219 (2013).
4. O'Regan, B. C., Durrant, J. R., Sommeling, P. M. & Bakker N. J. *J. Phys. Chem. C* **111,** 14001–14010 (2007).
5. Yang, H. G. *et al. Nature* **453,** 638–641 (2008).
6. Lee, M. M., Teuscher, J., Miyasaka, T., Murakami, T. N. & Snaith, H. J. *Science* **338,** 643–647 (2012).

EPIGENOMICS

Methylation's mark on inheritance

Epigenetic changes to the genome can have heritable effects. An epigenome-wide study of wild plants identifies shared patterns of such modifications and their associations with genetic information. SEE ARTICLE P.193

STEVEN EICHTEN & JUSTIN BOREVITZ

We have long known that genes carry the blueprint of inherited instructions that are passed from individuals to their offspring. More recent developments have shown that the make-up of an individual's epigenome — chemical and structural modifications to the genome that do not change the DNA sequence — may also contribute to this instruction set. One chemical modification, DNA methylation, is thought to target certain genetic elements, but it remains unclear what proportion of these marks is completely separate from genetic information. Reporting on page 193 of this issue, Schmitz *et al.*[1] begin to disentangle *de novo* methylation from that linked to variation in DNA sequence by studying the genome and epigenome of the model plant *Arabidopsis thaliana**.

The addition of a methyl group to cytosine residues in DNA is one of the most common epigenetic modifications in both plants and animals. Such methylation is frequently seen in repetitive sequences, is known to repress the movement of transposable genetic elements and, in some cases, can inhibit gene expression. There are three classes of methylation marks in plants: CG, CHG and CHH, in which C is methylated cytosine and H is any nucleotide residue other than guanine (G). These methylation classes differ in their mechanisms of establishment, maintenance and inheritance; and they have distinct roles in genomic regulation, thereby leading to differing effects on the characteristics — or phenotype — of an organism[2]. Previous analyses of genome-wide methylation profiles have identified many regions of variable DNA methylation in several plant systems[3,4].

"Schmitz and colleagues' study is an example of how association studies that include the epigenome can help to uncover additional sources of heritability."

Schmitz and colleagues took these analyses to the population level by studying genotypically distinct wild *A. thaliana* individuals collected from around the Northern Hemisphere. The authors obtained whole-genome DNA sequences for 217 individual plants, and, of these, whole-genome methylation profiles for 152 and gene-transcript profiles for 144 plants. From these data, they identified hundreds of thousands of sites at which a single nucleotide was methylated in some of the individuals (single methylation polymorphisms, or SMPs) and tens of thousands of differentially methylated regions (DMRs). More than 30% of these DMRs were directly associated with local sites of DNA-sequence variation. These associations define regions at which the genetic and epigenetic variants are inherited together. The authors also demonstrate that, as expected, many of these sites of association correlate with differential gene expression.

Using statistical tests that are designed to uncover associations between epigenetic variation and phenotype, Schmitz *et al.* were able to identify quantitative-trait loci (genomic regions that control trait variation) that were associated with differential methylation of all three classes. Furthermore, their examination of loci for which methylation was present across the samples led to the discovery that some genes have co-opted epigenetic regulatory mechanisms that were thought to target only transposable elements and repeat sequences. Many of these genes are important for the generation of gametes (sperm and egg cells or, in plants, pollen and ovules). It seems that DNA methylation prevents the expression of these genes during the vegetative phase of

*This article and the paper under discussion[1] were published online on 6 March 2013.

© 2013 Macmillan Publishers Limited. All rights reserved

plant development, but does not repress their expression in certain gametes because some components of the epigenetic machinery are inactive in these cells.

There are two extreme possibilities as to the potential relevance of heritable epigenetic information. The first is that epigenetic information is fully controlled by the genetic content of an individual. In this case, epigenetic profiling would be largely redundant because any epigenetic status is correlated with the individual's genotype. In such a scenario, it is also difficult to determine whether the genotype or the epigenotype is causal for any particular phenotype. In addition, the levels of epigenetic marks at unlinked sites may be distantly regulated by 'master loci' that also regulate complex phenotypes — Schmitz *et al.* identify examples of such cases.

The second case is that epigenetic variation forms within a uniform genetic background[5,6], allowing an additional level of heritable information to exist in an individual. These new 'epialleles' may be akin to single nucleotide polymorphisms (SNPs) in the genome that are rare in a population, but may be common in a family. However, SMPs and DMRs are reversible and can reoccur, whereas SNPs typically evolve only once and are stable. A high epiallelle frequency in a closely related population, as a result of inheritance and/or recurrent mutation, may explain phenotypic variation. In this case, an assessment of epigenetic information could add to our understanding of the heritable components that lead to an individual's traits. Early indications from inbred populations of *A. thaliana*, called synthetic epigenetic recombinant inbred lines[7,8], suggest this may be true. However, genetic variation may control the rate of formation of epialleles, returning us to the question of which form of variation is the causal level of control.

As in most cases, the reality is more complicated than either of the two possibilities alone. Epigenetic information seems to associate with the genetic content of an individual in a variety of ways[2], which means that we still do not know how often epigenetic states are associated with phenotypic variation. But Schmitz and colleagues' study is an example of how association studies that include the epigenome can help to uncover additional sources of heritability. Future epigenome-wide association studies should be large — involving thousands of individuals — and should aim to minimize genetic variation and allow epialleles to be common in the study sample, by targeting families or inbred lines that are separated by space and time. Epigenetic variation may turn out to be a transitory state in which phenotypic variation can be released, allowing rapid responses to environmental change. With time, genetic mutations can arise that then 'fix' this new state in a genetic, rather than an epigenetic, context. ■

Steven Eichten *is at the Microbial and Plant Genomics Institute, University of Minnesota, Saint Paul, Minnesota 55108, USA.*
Justin Borevitz *is in the Research School of Biology, Australian National University, Canberra, ACT 0200, Australia.*
e-mail: justin.borevitz@anu.edu.au

1. Schmitz, R. J. *et al. Nature* **495,** 193–198 (2013).
2. Richards, E. J. *Nature Rev. Genet.* **7,** 395–401 (2006).
3. Lister, R. *et al. Cell* **133,** 523–536 (2008).
4. Eichten, S. R. *et al. PLoS Genet.* **7,** e1002372 (2011).
5. Schmitz, R. J. *et al. Science* **334,** 369–373 (2011).
6. Becker, C. *et al. Nature* **480,** 245–249 (2011).
7. Johannes, F. *et al. PLoS Genet.* **5,** e1000530 (2009).
8. Reinders, J. *et al. Genes Dev.* **23,** 939–950 (2009).

STEM CELLS

Anatomy of an ovarian cancer

Whether ovarian cancer originates in the ovary or the surrounding tissues is a focus of debate. Work in mice now shows that stem cells that replenish the ovarian surface epithelium can be the initiators of this cancer. SEE LETTER P.241

JAMES D. BRENTON & JOHN STINGL

Stem cells in adult tissues are thought to be the cells of origin of cancers: these cells are long-lived, can self-renew and are susceptible to accumulating the minimal number of mutations that are necessary for carcinogenesis[1]. But the putative stem-cell population that gives rise to ovarian cancers has remained elusive. On page 241 of this issue, Flesken-Nikitin *et al.*[2] identify such cells in mice*.

The authors show that the cells covering the surface of the ovaries — the ovarian surface epithelium (OSE) — arise from a localized pool of stem cells that are clustered in the ovarian hilum. This is the region where blood vessels enter the ovary, and it forms the junction between the OSE and other cell types of the surrounding supporting structures, such as the ovarian bursa. Flesken-Nikitin *et al.* find that OSE stem cells express proteins such as Lgr5 and aldehyde dehydrogenase, which often characterize stem cells in other tissues[3,4].

Through elegant cell-lineage tracing experiments, the researchers also demonstrate that these cells can regenerate the OSE. Moreover, when mutated by deletion of the tumour-suppressor genes *Trp53* and *Rb1*, the stem cells — but not fully differentiated OSE cells — could efficiently generate tumours that resembled human ovarian cancer. The induced ovarian tumours expressed aldehyde dehydrogenase-1 and other markers typical of human ovarian cancers, suggesting that this mouse

*This article and the paper under discussion[2] were published online on 6 March 2013.

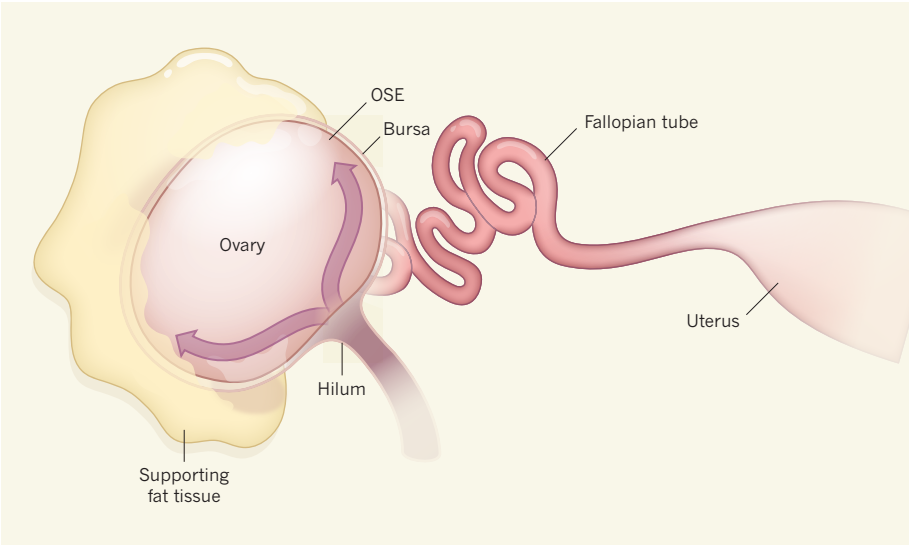


Figure 1 | Stem cells at the border between organs. The Fallopian tube is a conduit between the ovary and the uterus. Flesken-Nikitin *et al.*[2] show that, in mouse ovaries, stem cells (purple arrows) originating from the hilum region regenerate the ovarian surface epithelium (OSE) and, when mutated, can result in the formation of ovarian tumours.

© 2013 Macmillan Publishers Limited. All rights reserved

model reflects the human disease[5,6].

The finding that OSE stem cells, when mutated, can give rise to ovarian tumours is likely to reignite a long-standing debate about the cells of origin of high-grade serous carcinoma, the most common and aggressive type of epithelial ovarian cancer in humans.

Ovarian cancers were initially thought to originate in the ovaries. Large, fluid-filled malignant tumours are commonly found on, or replacing, the ovarian tissue in women undergoing surgery for ovarian cancer. However, it has been difficult to convincingly demonstrate early cancerous changes in the OSE. Healthy ovaries may contain small fluid-containing cysts, termed inclusion cysts, which are lined by OSE. Normally, these OSE cells have a roughly cuboidal shape, but in a few cases they show characteristics of dysplasia (abnormal accumulation of partially differentiated cells) and accumulation of the tumour-suppressor protein p53. Although the changes in OSE morphology could be an early precancerous change, unambiguous changes typical of high-grade serous carcinoma are observed only rarely in inclusion cysts (reviewed in ref. 7).

An ovarian origin for ovarian cancers has been challenged by research suggesting that these cancers arise in the epithelium that lines the Fallopian tubes (which are a conduit between the ovaries and the uterus) and in the uterus[8]. A Fallopian-tube origin is supported by pathological studies[9,10] showing that early invasive lesions can be found at the distal end of the tube, close to the ovaries. In addition, other subtypes of ovarian cancer (endometrioid and clear-cell cancer) can arise from endometriosis, a benign condition in which the endometrial lining of the uterus seeds on to the surface of other pelvic organs. The finding that the same tumour-suppressor genes and oncogenes are mutated in the endometriotic tissue adjacent to clear-cell ovarian cancers is strong evidence that endometriosis provides the cells of origin for these cancers[11,12].

Attractive though it might be to seek a reductionist model, neither of these models can account for all forms of human ovarian cancer. For example, occasionally, high-grade serous carcinomas in the ovaries show no signs of cancer in the Fallopian tube or elsewhere. The molecular explanations for the cell of origin are therefore likely to be more complex, and that is why Flesken-Nikitin and colleagues' finding that the transformation of OSE stem cells can recapitulate features of high-grade serous carcinoma is particularly intriguing.

At first glance, it is perplexing that tumours resembling those in the Fallopian tube can arise in the ovary. Indeed, OSE has a relatively featureless morphology, and it generally does not resemble the cells that constitute the bulk of ovarian tumours, which tend to have a more differentiated appearance and to show characteristics of normal Fallopian and uterine tissues.

One explanation for this paradox is that the OSE and the Fallopian tubes have a common early-embryological origin and that the OSE retains the potential to undergo differentiation into Fallopian-tube cells under the conditions prevalent during tumour progression[7] (Fig. 1). If so, then a revised explanation would be that the initial steps of carcinogenesis depend on reprogramming of the OSE into another epithelial-cell lineage. Conceptually, OSE stem cells could be more susceptible to this reprogramming because of their undifferentiated state.

Bear in mind that Flesken-Nikitin and co-authors' work does not exclude the possibility that epithelial ovarian cancers can also originate from non-stem cells. Indeed, it may be that a different set of oncogenic mutations induces cancers to originate from more-differentiated OSE cells. Studies in other epithelial tissues, such as the mammary gland, demonstrate[13,14] that oncogenic abnormalities, including loss of the *BRCA1* gene and increased activity of the Notch signalling cascade, can exert their effects on cells that are thought not to have stem-cell properties. Furthermore, increasing evidence suggests[15] that non-stem cells in both solid and blood-related tumours can acquire self-renewal properties and give rise to cancer.

Given the anatomical differences between mice and humans, the key questions now are whether similar localized populations of OSE stem cells exist in the human ovary and whether they are prone to malignant transformation. The answers may have a significant impact on the development of preventive strategies for ovarian cancer. ■

James D. Brenton *and* **John Stingl** *are at the Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE, UK.*
e-mail: james.brenton@cruk.cam.ac.uk

1. Sell, S. *Am. J. Pathol.* **176,** 2584–2494 (2010).
2. Flesken-Nikitin, A. *et al. Nature* **495,** 241–245 (2013).
3. Leushacke, M. & Barker, N. *Oncogene* **31,** 3009–3022 (2012).
4. Ma, I. & Allan, A. L. *Stem Cell Rev.* **7,** 292–306 (2011).
5. Silva, I. A. *et al. Cancer Res.* **71,** 3991–4001 (2011).
6. Kryczek, I. *et al. Int. J. Cancer* **130,** 29–39 (2012).
7. Auersperg, N. *Front. Biosci. (Schol. Edn)* **5,** 709–719 (2013).
8. Dubeau, L. *Gynecol. Oncol.* **72,** 437–442 (1999).
9. Piek, J. M. *et al. J. Pathol.* **195,** 451–456 (2001).
10. Lee, Y. *et al. J. Pathol.* **211,** 26–35 (2007).
11. Wiegand, K. C. *et al. N. Engl. J. Med.* **363,** 1532–1543 (2010).
12. Yamamoto, S. *et al. J. Pathol.* **225,** 189–194 (2011).
13. Lim, E. *et al. Nature Med.* **15,** 907–913 (2009).
14. Bouras, T. *et al. Cell Stem Cell* **3,** 429–441 (2008).
15. Visvader, J. E. *Nature* **469,** 314–322 (2011).

MEDICAL IMAGING

Sleuthing tissue fingerprints

A type of data-acquisition sequence in magnetic resonance imaging has been developed that rapidly and robustly quantifies properties of imaged tissue by elucidating a characteristic signal fingerprint. **SEE ARTICLE P.187**

E. BRIAN WELCH

A conventional magnetic resonance imaging (MRI) scan is designed to create contrast between tissue types. It does so on the basis of various physical properties that are related to the microenvironment of water molecules in the tissues. The resulting images are typically qualitative and merely represent a contrast that is weighted towards one tissue property or another. Although such qualitative images certainly contain clinically useful information, MRI scientists continue to search for methods that can efficiently quantify tissue properties. On page 187 of this issue, Ma *et al.*[1] present a kind of MRI data-acquisition sequence and image-formation algorithm that avoids the standard approach of acquiring an image weighted towards a single tissue property. This innovative method is based on the collection of magnetic resonance 'fingerprints' that reflect the combined effect of multiple underlying tissue properties. The approach has the potential to make quantitative mapping of tissue properties feasible in the clinical setting and to lead to improved diagnoses.

The pattern of curves on the human fingertip does not directly encode specific information such as an individual's name or sex, but its uniqueness does provide a way to recognize an individual and so link it to known attributes for that person. In other contexts, the test fingerprint can be used to describe any unique and recognizable signal that can be linked back to a specific set of properties. Conventional MRI methods that create quantitative maps of

© 2013 Macmillan Publishers Limited. All rights reserved

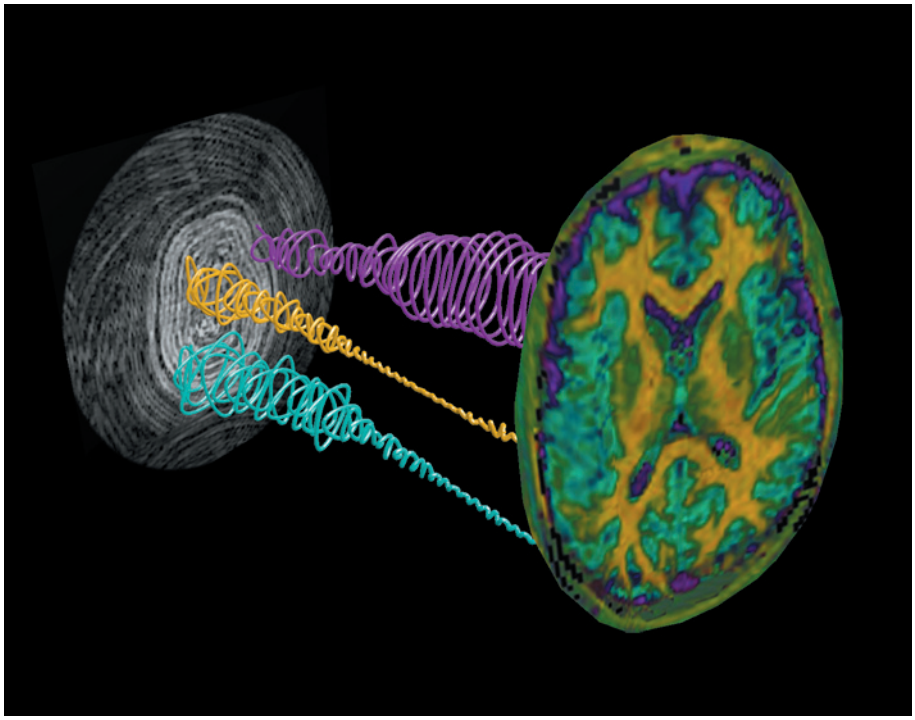

DAN MA & MARK GRISWOLD

Figure 1 | Magnetic resonance fingerprinting. Ma and colleagues' fingerprinting approach[1] for magnetic resonance imaging (MRI) converts what seem to be random signals (coloured lines) into coherent maps of MRI properties of the tissue being imaged, in this case a human brain. The black-and-white image in the background displays a single frame from a magnetic resonance fingerprinting acquisition. The magnetic resonance fingerprinting data are combined to create the false-colour quantitative map of tissue fraction in the foreground. Yellow, white matter; cyan, grey matter; magenta, cerebrospinal fluid; other colours represent mixtures of these basic tissues.

tissue properties are time consuming to collect and susceptible to corruption if the subject moves. The magnetic resonance fingerprinting approach described by Ma *et al*. is both fast and resilient to inconsistencies in the acquired data. The strength of the method comes from the use of a specialized data-acquisition sequence that elucidates a characteristic signal, or fingerprint, for a given spatial tissue location. That signal can then be compared with a set of precalculated fingerprint signals that are associated with known tissue properties. Once the fingerprint for a given spatial location has been recognized, that same location can be assigned the associated tissue properties in quantitative maps (Fig. 1).

In medical imaging, a scanner collects signals of emitted or transmitted energy after some type of input energy — such as that of X-ray photons, ultrasonic waves and radio-frequency waves — interacts with the object being imaged, for example the organs and tissues of the human body. In general, the physical properties of the tissue and the microenvironment surrounding it determine the intensity of the collected signals. The signals also depend on the selection of data-acquisition parameters made by the scanner operator. To form an image with an MRI scanner, a powerful magnet is used to generate a strong magnetic field that affects the magnetic spins of protons in the water molecules prevalent throughout the body. A small, but detectable, majority of the magnetic spins align themselves with the strong magnetic field. The spins precess (rotate) about the strong magnetic field at a characteristic frequency known as the Larmor frequency[2], much like a gyroscope precesses about the force vector created by gravity. The precessing magnetic spins can be excited from an equilibrium energy state to a state of higher energy by using a radio-frequency wave that has a frequency matching the Larmor frequency. Once the input radio-frequency energy is deactivated, the excited spins return to the equilibrium energy state and emit radio-frequency energy, which can be detected to form an image.

The intensity of the emitted radio-frequency energy is a function of several tissue properties, including the proton density (*PD*) at a given spatial location and how quickly the excited spins return (relax) to the equilibrium energy state. The rate of relaxation is described by two time constants[3] (T_1 and T_2) that reflect relaxation in the geometric planes respectively parallel and perpendicular to the strong magnetic field. The signal intensity for a given spatial location in almost all magnetic resonance images can be defined as a function of *PD*, T_1, T_2 and a few data-acquisition sequence parameters. Obtaining maps of *PD*, T_1 and T_2 is important in a clinical setting because pathological changes to tissue often alter one or more of the values for these three parameters. The importance of these tissue properties is apparent when one considers that most conventional clinical magnetic resonance images are weighted towards *PD*, T_1 or T_2 to reveal pathological conditions.

Since the introduction of MRI in the early 1970s[4], magnetic resonance physicists and engineers have continued to develop approaches to show information about the tissues and organs being scanned. The clinical utility of rapidly and robustly mapping *PD*, T_1 and T_2 makes the technique described by Ma *et al*. extremely valuable. The authors use a customized version of a data-acquisition protocol known as an inversion-recovery balanced steady-state free-precession (IR-bSSFP) sequence that is sensitive to all the tissue properties of interest[5]. Unlike the conventional approach of keeping sequence parameters constant throughout the repeated data readouts of the IR-bSSFP protocol, the authors pseudorandomly varied the amplitude of radio-frequency excitation and time spacing between repeated excitations to cause the collected MRI signal at a given spatial location to behave in a pseudorandom manner. However, the pseudorandom MRI-signal response pattern is actually a function of *PD*, T_1 and T_2 at that spatial location. The authors refer to this signal pattern as the magnetic resonance fingerprint for a particular combination of tissue properties.

Ma and colleagues present compelling evidence for the success of the strategy using 'imaging phantoms' with known properties and for *in vivo* brain images. The method is robust in the presence of bulk motion of the object being imaged and in the case of data undersampling. The reduction of acquired data markedly decreases scan time by a factor of nearly 50-fold compared with the time of the fully sampled scan. The approach should be adaptable to other MRI data-acquisition sequences, and could be extended to include the effects of other tissue and microenvironment properties that are relevant to the adapted MRI sequence types. Perhaps the most exciting possible future application of the method is the collection of magnetic resonance fingerprints in patients with confirmed pathology, such as brain tumours, that can be difficult to grade and classify. A library of fingerprints that are correlated to known pathology could be used to improve diagnosis for such patients. ■

E. Brian Welch *is at the Vanderbilt University Institute of Imaging Science, Nashville, Tennessee 37232–2310, USA.*
e-mail: brian.welch@vanderbilt.edu

1. Ma, D. *et al. Nature* **495,** 187–192 (2013).
2. Kellogg, J., Rabi, I., Ramsey, N. & Zacharias, J. *Phys. Rev.* **56,** 728–743 (1939).
3. Levitt, M. H. *Spin Dynamics* 2nd edn (Wiley, 2008).
4. Lauterbur, P. C. *Nature* **242,** 190–191 (1973).
5. Schmitt, P. *et al. Magn. Reson. Med.* **55,** 177–186 (2006).

© 2013 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature11971

Magnetic resonance fingerprinting

Dan Ma[1], Vikas Gulani[1,2], Nicole Seiberlich[1], Kecheng Liu[3], Jeffrey L. Sunshine[2], Jeffrey L. Duerk[1,2] & Mark A. Griswold[1,2]

Magnetic resonance is an exceptionally powerful and versatile measurement technique. The basic structure of a magnetic resonance experiment has remained largely unchanged for almost 50 years, being mainly restricted to the qualitative probing of only a limited set of the properties that can in principle be accessed by this technique. Here we introduce an approach to data acquisition, post-processing and visualization—which we term 'magnetic resonance fingerprinting' (MRF)—that permits the simultaneous non-invasive quantification of multiple important properties of a material or tissue. MRF thus provides an alternative way to quantitatively detect and analyse complex changes that can represent physical alterations of a substance or early indicators of disease. MRF can also be used to identify the presence of a specific target material or tissue, which will increase the sensitivity, specificity and speed of a magnetic resonance study, and potentially lead to new diagnostic testing methodologies. When paired with an appropriate pattern-recognition algorithm, MRF inherently suppresses measurement errors and can thus improve measurement accuracy.

Magnetic resonance techniques such as NMR spectroscopy and magnetic resonance imaging (MRI) are widely used throughout physics, biology and medicine because of their ability to generate detailed information about numerous important material or tissue properties, including those reflective of many common disease states[1–4]. However, in practice magnetic resonance acquisitions are often restricted to a qualitative or 'weighted' measurement of a limited set of these properties; the magnetic resonance signal intensity is almost never quantitative by itself. The same material can have different intensities in different data sets depending on many factors, including the type and set-up of the scanner, the detectors used, and so on. Because of this, the quantitative analysis of magnetic resonance results typically focuses on differences between spectral peaks, spatial locations or different points in time. Even in clinical MRI today, a tissue or material is typically referred to as being 'hyperintense' or 'hypointense' compared to another area, which may not provide a quantitative indication of the severity of the differences, and may have reduced sensitivity to global changes. Thus robust, fully quantitative multiparametric acquisition has long been the goal of research in magnetic resonance[5–8]. However, the quantitative methods developed to date typically provide information on a single parameter at a time, require significant scan time, and are often highly sensitive to system imperfections. Simultaneous, multiparametric measurements are almost always impractical owing to scan time limits and a high sensitivity to the measurement set-up and experimental conditions. Thus purely qualitative magnetic resonance measurements remain the standard today, particularly in clinical MRI.

Here we introduce a novel approach, namely MRF, that may overcome these constraints by taking a completely different approach to data acquisition, post-processing and visualization. Instead of using a repeated, serial acquisition of data for the characterization of individual parameters of interest, MRF uses a pseudorandomized acquisition that causes the signals from different materials or tissues to have a unique signal evolution or 'fingerprint' that is simultaneously a function of the multiple material properties under investigation. The processing after acquisition involves a pattern recognition algorithm to match the fingerprints to a predefined dictionary of predicted signal evolutions. These can then be translated into quantitative maps of the magnetic parameters of interest.

MRF is related to the concept of compressed sensing[9–12], and shares many of its predicted benefits. For example, preliminary results show that MRF could acquire fully quantitative results in a time comparable to a traditional qualitative magnetic resonance scan, without the high sensitivity to measurement errors found in many other fast methods. Most importantly, MRF has the potential to quantitatively examine many magnetic resonance parameters simultaneously given enough scan time, whereas current magnetic resonance techniques can only examine a limited set of parameters at once. Thus MRF opens the door to computer-aided multiparametric magnetic resonance analyses, similar to genomic or proteomic analyses, that could detect important but complex changes across a large number of magnetic resonance parameters simultaneously. When an appropriate pattern recognition algorithm is used, MRF also provides a new and more robust behaviour in the presence of noise or other acquisition errors that may lead to the near complete suppression of deleterious effects stemming from these factors. Although we focus on demonstrating the feasibility for MRI in this study, it is rather straightforward to translate these results to other magnetic resonance fields, such as multiparametric NMR spectroscopy, dynamic contrast enhanced MRI and dynamic susceptibility contrast MRI[13].

Generation and recognition of MRF signals

The key assumption underlying the MRF concept is that unique signal evolutions, or fingerprints, can be generated for different materials or tissues using an appropriate acquisition scheme. Here we demonstrate that this is possible through the continuous variation of the acquisition parameters throughout the data collection. Variations in the pulse sequence parameters during acquisition have been used previously in MRI and magnetic resonance spectroscopy to reduce the signal oscillations[14] and to improve the spectral response[15–17]. However, these variations were primarily used in a preparation phase or to make the signal more constant. Randomized sampling patterns have also been used previously to aid in the separation of spatiotemporal signals in moving objects or substances with different resonance frequencies[18–20]. Here we demonstrate that temporal and spatial incoherence required in MRF can be achieved by varying acquisition parameters—such as the flip angle and phase of radio frequency pulses, the repetition time, echo time and sampling patterns—in a pseudorandom manner.

[1]Department of Biomedical Engineering, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA. [2]Department of Radiology, Case Western Reserve University and University Hospitals of Cleveland, 11100 Euclid Avenue, Cleveland, Ohio 44106, USA. [3]Siemens Healthcare USA, 51 Valley Stream Parkway, Malvern, Pennsylvania 19355, USA.

©2013 Macmillan Publishers Limited. All rights reserved

After the data are acquired, the separation of the signal into different material or tissue types can be achieved through pattern recognition. In its simplest form, this process is analogous to matching a person's real fingerprint to a database: once a match is made, a host of additional information about the person, such as name, address and phone number, can be obtained simultaneously once the fingerprint sample is identified. In MRF, this pattern recognition can take place through many means. In the current implementation, we construct a dictionary that contains signal evolutions from all foreseeable combinations of materials and system-related parameters—for example, the longitudinal relaxation time, T_1, the transverse relaxation time, T_2 and off-resonance frequency are included in this study. Other properties could also be measured, such as diffusion and magnetization transfer using the well-established Bloch equation formalism of magnetic resonance [21,22]. Once this dictionary of possible signal evolutions is generated, a matching or pattern recognition algorithm[23,24] is then used to select a signal vector or a weighted set of signal vectors from the dictionary that best correspond to the observed signal evolution. All the parameters that were used to build this signal vector in the dictionary can then be retrieved simultaneously. At present, the calculation of a complete dictionary containing the realistic range of T_1, T_2 and off-resonance frequency requires only a few minutes on a modern desktop computer.

It should be noted that there are near-infinite possibilities for MRF-compatible pulse sequences. Other magnetic resonance parameters of interest can be investigated by identifying pulse sequence components that impart differential sensitivity to the parameters of interest. Moreover, different components can be varied simultaneously, adding the potential for a highly efficient experimental design that allows almost any material characteristic visible using magnetic resonance to be analysed in a quantitative way using MRF.

Validation of the concept

For a proof-of-principle implementation, an MRF acquisition based on an inversion-recovery balanced steady state free-precession (IR-bSSFP) sequence was used (Fig. 1a). This choice of this basic pulse sequence was based on the extensive existing knowledge about IR-bSSFP signal evolution, and its sensitivity to T_1, T_2 and off-resonance frequency[25]. After each radio-frequency pulse, one interleaf of a variable density spiral (VDS) read out[26] was acquired, as shown in Fig. 1b. Such a VDS trajectory has been used in fast imaging[27] and for the reduction of undersampling errors[28]. Two MRF acquisition patterns with randomized flip angle and repetition time were used as shown in Fig. 1c and d in separate scans to demonstrate the flexibility of the choice of the acquisition parameters.

Figure 2a and b show the simulated signal evolution curves that would be expected from four commonly encountered tissues of the brain (fat, white matter, grey matter and cerebrospinal fluid) using the schematic implementation shown in Fig. 1c and d, respectively. Each tissue type has characteristic T_1 and T_2 values and thus each signal evolution has a different shape, which confirms that it is possible to satisfy this fundamental assumption in MRF. Note also that the signal levels in these evolutions represent a large fraction of the equilibrium magnetization (which is normalized to one in these figures.) Conventional spoiled steady-state sequences typically generate signal levels corresponding to 1–10% of the equilibrium magnetization. Figure 2c and d show an acquired signal evolution curve from fully sampled experiments on manufactured agar 'phantoms' and its match to the dictionary by using the acquisition pattern shown in Fig. 1c and d, along with the recovered T_1, T_2, proton density (M_0) and off-resonance frequency values. MRF was able to match the signal to the corresponding dictionary entry and obtain the same T_1 and T_2 values from both sequence patterns. A video of the signal evolution from a fully sampled *in vivo* scan is available (Supplementary Video 1), demonstrating the oscillating nature of the MRF signal observed *in vivo*.

Accelerated MRF acquisitions

In addition to simultaneously quantifying multiple parameters, the error tolerance of MRF can be significantly better than that of conventional MRI. Because MRF is based on pattern recognition in a setting where the form of all predicted signal evolutions is known, MRF should be less sensitive to errors during the measurement. This is similar to conventional fingerprint recognition techniques, which often contend with smudges and partial fingerprint information. In particular, the interaction of the temporal and spatial incoherence possible in MRF provides new opportunities to accelerate image acquisition through rejection of spatial undersampling errors. In order to test the limits of this acceleration, the same MRF sequence as shown in Fig. 1a–c was modified to use only one spiral readout in each acquisition block. Therefore, the data collected are only 1/48th of the normally required data at each time point, resulting in a total acquisition time of 12.3 s, corresponding to 1,000 sampled time points. (See Fig. 3a and Supplementary Video 2.) The signal evolutions from all 1,000 undersampled time points were used directly to match one entry from the dictionary to quantify T_1, T_2, M_0 and

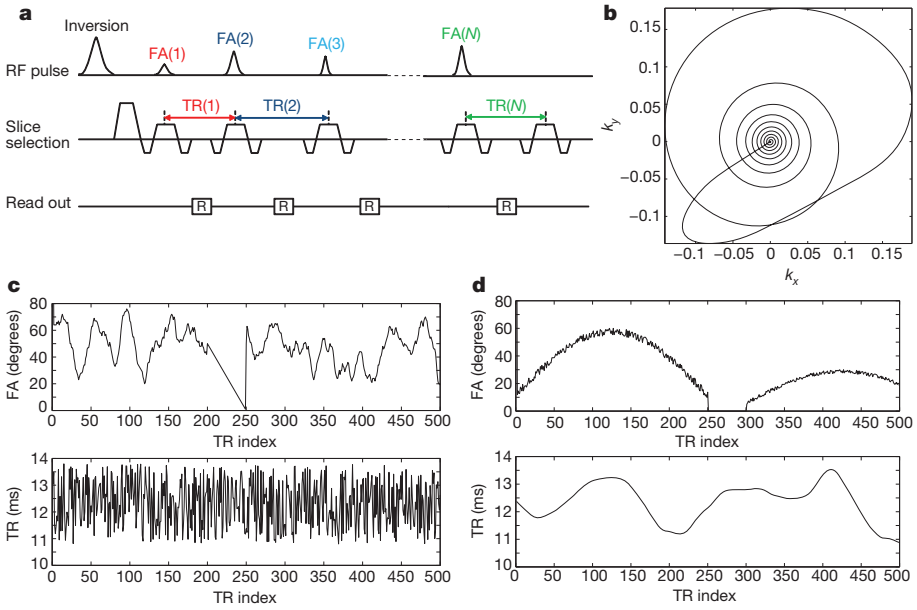


Figure 1 | MRF sequence pattern. **a**, Acquisition sequence diagram. In each subsequent acquisition block, identified by a repetition time index (TR index; TR(1)...TR(*N*)), various sequence components are varied in a pseudorandom pattern. FA, flip angle. **b**, Here, one variable density spiral trajectory was used per repetition time. The Fourier coefficients sampled by the variable density spiral trajectory (given by the coordinates k_x and k_y) are rotated from one repetition time to the next. **c**, **d**, Examples of the first 500 points of flip angle and repetition time patterns that were used in this study.

©2013 Macmillan Publishers Limited. All rights reserved

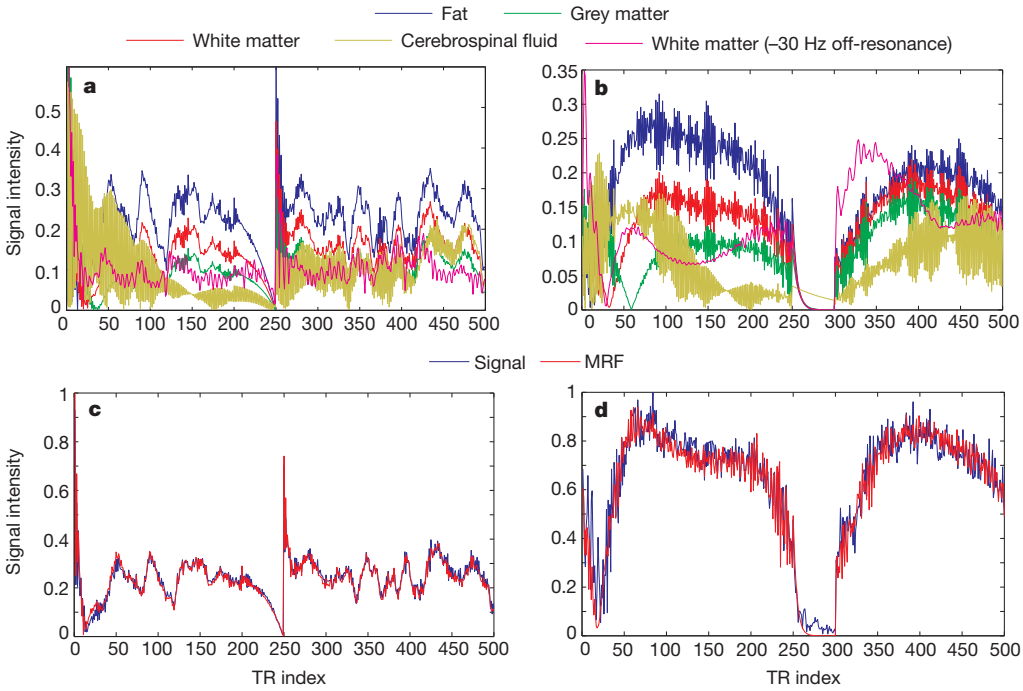


Figure 2 | Signal properties and matching results from phantom study. **a**, **b**, Simulated signal evolution curves corresponding to four normal brain tissues using the sequence patterns in Fig. 1c and d, respectively, as a fraction of the equilibrium magnetization. The curve from white matter with off-resonance is also plotted. **c**, **d**, Measured signal evolutions from one of eight phantoms using different sequence patterns and their dictionary match. The estimated T_1, T_2 and off-resonance frequencies are 340 ms, 50 ms and −4 Hz (**c**) and 340 ms, 50 ms, −13 Hz (**d**). The plots are normalized to their maximum value.

off-resonance simultaneously, as shown in Fig. 3b. Because these errors are incoherent with the expected MRF signals, they are largely ignored by the following processing steps. Figure 3c–f shows that high quality estimates of the magnetic resonance parameters are generated even with this significant level of undersampling. White-matter, grey-matter and cerebrospinal-fluid regions were then selected from the resultant maps. The mean T_1 and T_2 values obtained from each region are listed in Table 1 and are within the range of previously reported literature values[29–32]. The shortened T_2 value in CSF is probably due to out-of-plane flow in this two-dimensional experiment. A similar effect can also be observed in conventional T_2 mapping techniques[33]. We also note that the roughly −220 Hz chemical shift of fat protons is clearly visualized in the off-resonance map.

Motion error tolerance in MRF

Because motion is one of the most common sources of error in an MRI scan, a motion-corrupted scan was performed using the accelerated MRF acquisition described in the previous section. The subject was instructed to randomly move his head for the last 3 s of a total 15-s scan. Supplementary Video 3 shows the random motion as well as severe undersampling artefacts in the reconstructed images. Figure 4 compares the quantitative maps from the data with and without the motion-corrupted data. The maps acquired during motion show almost no sensitivity to the motion, and show nearly the same quality and anatomy as the maps from the motion-free data, thus indicating that the signal changes resulting from motion were uncorrelated with the evolutions included in the dictionary, and were largely ignored by the pattern recognition algorithm.

Accuracy and efficiency of MRF

The accuracy and efficiency of the MRF acquisitions were compared with alternative mapping strategies, namely, standard spin-echo sequences[34] as well as modern rapid combined T_1 and T_2 mapping methods DESPOT1 and DESPOT2 (driven equilibrium single pulse observation of T_1 and T_2, respectively)[30] using manufactured agar phantoms. Figure 5a compares the phantom T_1 and T_2 values from these methods. The concordance coefficient correlations for T_1 and T_2 between MRF and spin-echo sequence were 0.988 and 0.974, respectively. The concordance coefficient correlations for T_1 and T_2 between DESPOT and spin-echo sequence were 0.956 and 0.914, respectively. The high concordance correlation coefficients indicate that both methods are in good agreement with standard spin-echo measurements, and that MRF shows a better accuracy than DESPOT1 and DESPOT2.

The theoretical comparison of the efficiency from various mapping methods has been presented[35,36] and is based on a measure of precision per square root of scan time. In those publications, DESPOT1 and DESPOT2 were shown to have greater efficiency than all previously known conventional and accelerated mapping strategies[36]. As can be seen in Fig. 5b, MRF outperforms both DESPOT1 and DESPOT2 by an average factor of 1.87 and 1.85, respectively. For example, at a T_1 of ~1,280 ms, MRF shows an average efficiency for estimation of T_1 of 24.2, whereas DESPOT1 has an average efficiency of 10.89. This means that for this T_1 value, MRF achieves a precision of ±15.2 ms (or 1.2%) in 12 s of scan time, whereas the precision in DESPOT1 would be ±33.9 ms (or 2.6%) for the same scan time. The DESPOT methods apparently display higher efficiency from the single phantom, with T_1 of 360 ms and T_2 of 53 ms. However, in this one particular phantom, DESPOT overestimated the values of T_1 and T_2 by 23% and 42%, respectively, compared to the standard values, as can be seen in Fig. 5a, thus causing an erroneous increase in the apparent efficiency. Note that these efficiency estimates do not include the waiting times between the acquisition of the different sub-sequences in DESPOT, nor do they include the time required to reach steady state during each acquisition, and thus should be viewed as conservative estimates of the performance of MRF when compared to DESPOT1 or DESPOT2.

Because there is no steady state in the signal evolution from MRF, new information will be continuously added by longer acquisitions. Figure 5c and d illustrates the changes of mean and standard deviation as different acquisition times were used to quantify T_1 and T_2, with a clear trend towards lower error at longer acquisition times. Thus one can select a trade-off between precision and scan time.

Discussion and conclusions

The MRF concept presented here is a new approach to magnetic resonance and provides many opportunities to extend such measurements beyond their current limits. This originates from the unique pulse sequence design concept in MRF, where the goal is to generate unique signal evolutions that can be matched to theoretical signal evolutions and subsequently yield underlying quantitative information

©2013 Macmillan Publishers Limited. All rights reserved

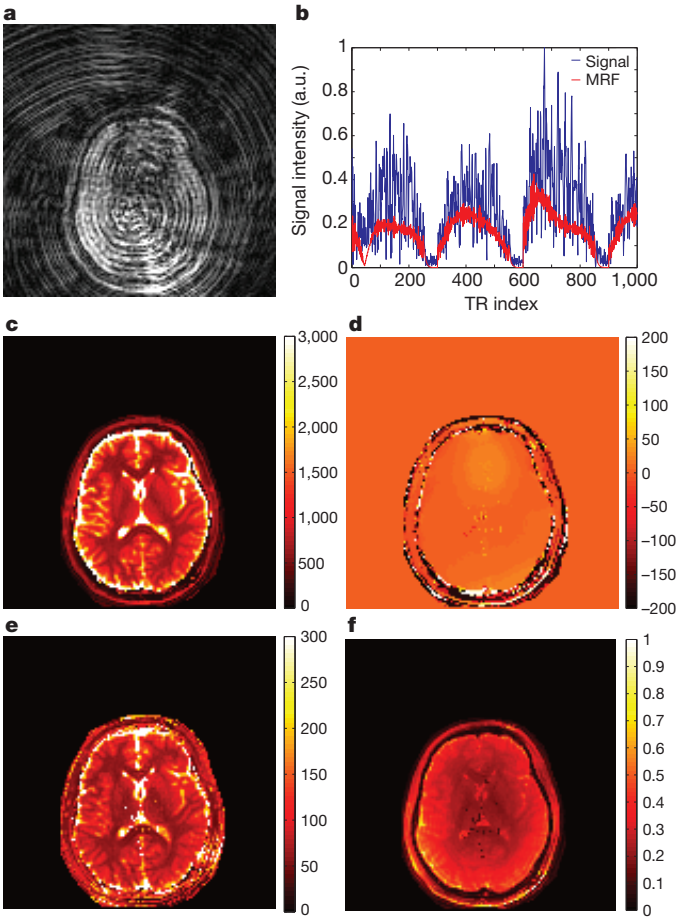


Figure 3 | MRF results from highly undersampled data. **a**, An image at the 5th repetition time out of 1,000 was reconstructed from only one spiral readout, demonstrating the significant errors from undersampling. **b**, One example of acquired single evolution and its match to the dictionary. Note the significant interference resulting from the undersampling. a.u., arbitrary units. **c**–**f**, The reconstructed parameter maps show a near complete rejection of these errors based solely on the incoherence between the underlying MRF signals and the undersampling errors; **c**, T_1 (colour scale, milliseconds); **d**, off-resonance frequency (colour scale, hertz); **e**, T_2 (colour scale, milliseconds); and **f**, proton density (M_0) (normalized colour scale). These data required 12.3 s to acquire.

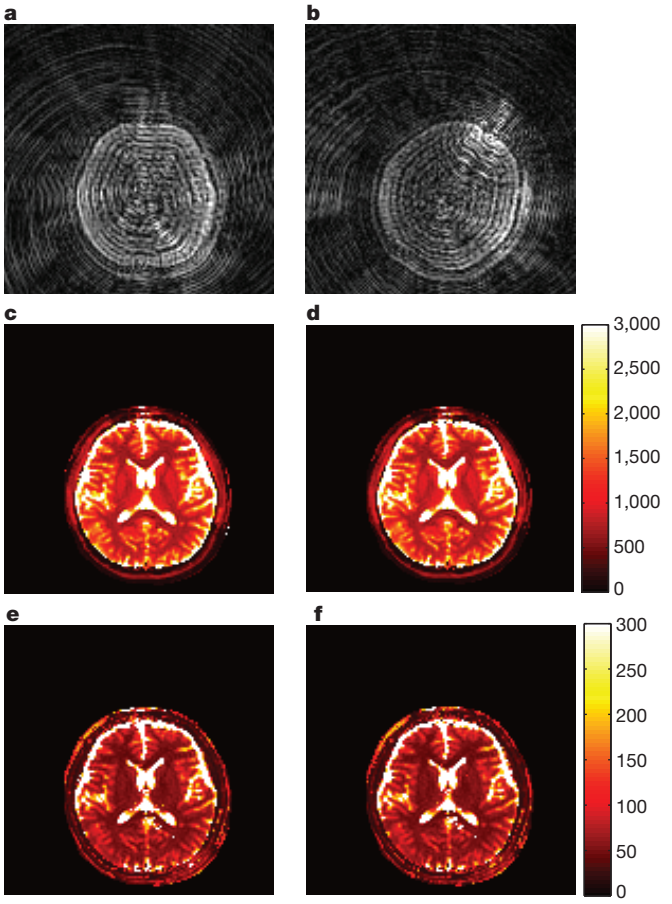


Figure 4 | Demonstration of error tolerance in the presence of motion. **a**, **b**, Reconstructed images acquired at the 12th second (**a**) and at the 15th second (**b**) demonstrate the large shift in the head position. **c**–**f**, The resulting MRF maps are nearly identical, demonstrating a rejection of both undersampling and motion errors that are uncorrelated with the expected signal evolution. **c**, **d**, T_1 (colour scale, milliseconds); **e**, **f**, T_2 (colour scale, milliseconds); **c** and **e** are from the first 12 s that has no motion, **d** and **f** are from entire 15 s that includes the motion.

about the material, tissue or pathology of interest. Because there is no *a priori* requirement on the shape of the signal evolution curves, there are more degrees of freedom in designing an MRF acquisition, where parameters such as repetition time, echo time, radio frequency pulses and sampling trajectories (among others) can be varied together to produce the simultaneous sensitivity to numerous tissue properties. The ability to analyse oscillating signals in MRF also provides the opportunity to use larger fractions of the available magnetization than methods that rely on a steady-state signal, which is a significant factor contributing to the higher efficiency in MRF. In addition, the oscillatory signal in MRF allows one to sample more informative points along a longer signal evolution as compared to conventional methods which always reach a steady state level after some finite amount of time. Specifically, our initial results here demonstrate that the efficiency of MRF is approximately 1.8 times higher than the DESPOT methods, which were previously the most efficient methods for the measurement of relaxation parameters. Thus the direct prediction of the oscillating, incoherent signal evolutions through the Bloch simulation provides us the potential to obtain new quantitative information that is impractical today because of the prohibitively long scan times required, especially in biological samples and patients.

Table 1 | *In vivo* data

	T_1 (ms)	T_2 (ms)
White matter (this work)	685 ± 33	65 ± 4
White matter (previously reported)	608–756	54–81
Grey matter (this work)	1,180 ± 104	97 ± 5.9
Grey matter (previously reported)	998–1,304	78–98
Cerebrospinal fluid (this work)	4,880 ± 379	550 ± 251
Cerebrospinal fluid (previously reported)	4,103–5,400	1,800–2.460

Shown are comparisons of MRF results and reference values[29–32] in different brain regions.

As demonstrated by the results shown here, MRF has the potential to significantly reduce the effects of errors during acquisition through its basis in pattern recognition. Acquisition errors may globally reduce the probability of a match of an observed signal to any given fingerprint, but as long as the errors do not cause another fingerprint to become the most likely match, the correct quantitative identification will still be made. Ideally, the sequence pattern will be designed so that the various fingerprints from different tissues and materials are as independent as possible, thus ensuring this robustness against motion and other practical errors.

Commercial magnetic resonance scanners include methods to minimize the effects of unavoidable system imperfections. However, these inaccuracies are becoming increasingly important as magnetic resonance technology is pushed to its limits, such as the use of very high magnetic fields or physically larger systems. MRF provides a route to model and account for system imperfections, such as inhomogeneities in both the static magnetic field (B_0) and the

©2013 Macmillan Publishers Limited. All rights reserved

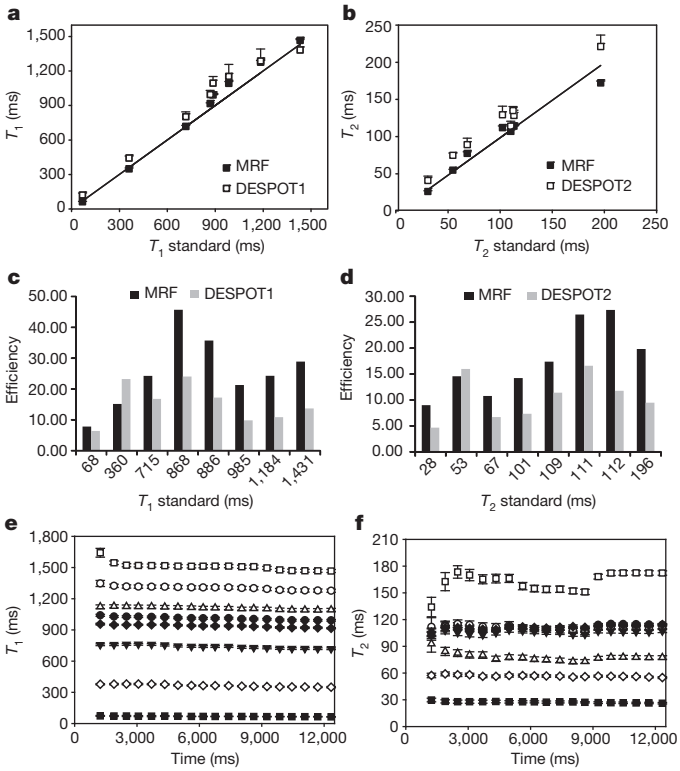


Figure 5 | Accuracy, efficiency and error estimation for MRF and DESPOT. **a**, **b**, The T_1 and T_2 values retrieved from MRF from eight phantoms are compared with those acquired from DESPOT1 (**a**), DESPOT2 (**b**) and a standard spin-echo sequence. **c**, **d**, The efficiency of MRF compared to DESPOT1 (**c**) and DESPOT2 (**d**) at different T_1 and T_2 values; efficiency is assessed as precision per (acquisition time)$^{1/2}$. MRF has an average of 1.87 and 1.85 times higher efficiency than DESPOT1 and DESPOT2, respectively. **e**, **f**, Obtained values of T_1 (**e**) and T_2 (**f**) as a function of acquisition time. Data in **a**–**d** show mean ± s.d. of the results over a 25-pixel region in the centre of each phantom, and are smaller than the symbols for most MRF results.

applied radio frequency field (B_1), by adding these parameters into the dictionary simulation. Because both MRF and DESPOT2 are based on a bSSFP sequence, which is known to be sensitive to field inhomogeneities[30,37], Supplementary Fig. 2 compares the T_2 maps acquired from MRF and DESPOT2 from an *in vivo* scan. Because off-resonance is not taken into account in the DESPOT2 model, the T_2 map from DESPOT2 shows areas of signal voids resulting from susceptibility effects at the air–tissue interfaces. MRF naturally incorporates these effects into the fingerprints, and thus the maps generated by MRF do not show these errors. Thus MRF could, for example, provide higher quality results using the current generation of magnetic resonance scanners. Alternatively, MRF could also allow the design of lower cost magnetic resonance scanners that can provide the same quality as today's high end systems through application of MRF models.

Because of its ability to provide quantitative results across many parameters simultaneously, MRF could lead to the direct identification of a material, tissue or pathology solely on the basis of its fingerprint. For example, many cancer cells show changes in multiple magnetic resonance parameters (for example, T_1, T_2 and self-diffusion tensor), a combination (though no single parameter) of which could potentially characterize them as different from all surrounding normal tissue types, and thus potentially separable. In an ideal situation, each given material, tissue or pathology would have its own signal evolution which would be orthogonal to all other signal evolutions. The MRF concept also implies that completely different acquisition schemes are possible in cases where one is only interested in the presence or absence of a particular material or disease state. For example, one could do a very rapid MRF scan of a bulk area of material or tissue and compare the measured signal evolutions against the set of known states of interest. This measurement could either indicate the presence of the material or disease of interest, or indicate its absence within a margin of error. This feature could result in very rapid and accurate screening procedures. In particular, this feature may help to relax the required spatial resolution of an MRI examination, thus increasing the speed, and potentially reducing the cost, of such an examination. A preliminary example of this kind of visualization is shown in Supplementary Information Section 3. Using the MRF concept, the operation of the magnetic resonance unit will also be greatly simplified, because the 'all in one' scan concept of MRF has the potential to reduce the dozens of parameters currently presented to the magnetic resonance operator to a simple 'scan' button.

It is important to note that the proof-of-principle implementation of MRF shown here is but one of the many possibilities that could be used for this technique, and both the sequence design/implementation and post-processing methods will continue to be a significant open area of research, just as sequence design has advanced over the decades since the conventional methods have been introduced. Other, more advanced pattern recognition algorithms[38–42] will probably improve the performance of MRF. For spatially encoded MRI applications, the parameters retrieved from MRF are far fewer than the number of pixels in the images, and because the signals generated are largely incoherent, MRF has the additional potential to be highly accelerated through combination with other compressed sensing methods for accelerated spatial encoding, in addition to the now standard parallel imaging methods[43–45], neither of which were included here. Any of these methods would reduce the undersampling errors seeing in Fig. 3a even before the pattern recognition step, which should result in higher quality results. We have recently published data indicating that we can achieve a reduction in imaging time of about ten times for a two-dimensional slice using parallel imaging alone[43,45,46]. Also, it should be noted that the proof-of-principle results shown here only take advantage of two spatial dimensions for undersampling, whereas it is well known that taking full advantage of undersampling in all three spatial dimensions gives higher performance than a two-dimensional acquisition owing to the reduced power of the resulting errors at any given undersampling factor[47]. Thus a combination of an optimized three-dimensional MRF pulse sequence with parallel imaging and more advanced pattern matching algorithms will allow realization MRF in very short scan times.

METHODS SUMMARY

Sequence design. After an initial inversion pulse, the sequence pattern shown in Fig. 1c used a pseudorandomized series (Perlin noise[48]) of flip angle and a random repetition time from 10.5 to 14 ms. The flip angle pattern in Fig. 1d contained repeating sinusoidal curves with a period of 250 acquisitions and alternating maximum flip angles. The repetition time was a Perlin noise pattern. The radio frequency phase for both of the patterns alternated between 0 and 180° on successive radio frequency pulses. The variable density spiral-out trajectory was designed using minimum-time gradient design[49].

Dictionary design. A total of 563,784 signal time courses, each with 1,000 time points, with different sets of characteristic parameters (T_1, T_2 and off-resonance) were simulated for the dictionary. One dictionary entry was selected for each measured pixel location using template matching. In this case, the vector dot-product was calculated between the measured time course and all dictionary entries using complex data for both. The dictionary entry with the highest dot-product was then selected as most likely to represent the true signal evolution. The M_0 value is then the multiplicative constant derived by fitting the acquired data to this dictionary entry.

Data acquisition. All data were acquired on a 1.5-T whole body scanner (Espree, Siemens Healthcare) with a standard 32-channel head receiver coil. Images from each acquisition block were reconstructed separately using non-uniform Fourier transform[50]. The resultant time series of images was used to determine the value for the parameters (T_1, T_2, M_0 and off-resonance) as described above.

Statistical analysis. Quantitative estimates of the errors and efficiencies of MRF, DESPOT1 and DESPOT2 were calculated pixel-wise using a bootstrapped Monte Carlo method[51]. The means and standard deviations of T_1 and T_2 along the 50

©2013 Macmillan Publishers Limited. All rights reserved

repetitions were calculated, and averaged within a square (5 pixel × 5 pixel) region of interest for each phantom. The concordance correlation coefficients and efficiency were calculated as in refs 30 and 52.

Full Methods and any associated references are available in the online version of the paper.

Received 11 September 2012; accepted 30 January 2013.

1. Bartzokis, G. *et al. In vivo* evaluation of brain iron in Alzheimer disease using magnetic resonance imaging. *Arch. Gen. Psychiatry* **57,** 47–53 (2000). .
2. Larsson, H. B. *et al.* Assessment of demyelination, edema, and gliosis by *in vivo* determination of T1 and T2 in the brain of patients with acute attack of multiple sclerosis. *Magn. Reson. Med.* **11,** 337–348 (1989).
3. Pitkänen, A. *et al.* Severity of hippocampal atrophy correlates with the prolongation of MRI T2 relaxation time in temporal lobe epilepsy but not in Alzheimer's disease. *Neurology* **46,** 1724–1730 (1996).
4. Williamson, P. *et al.* Frontal, temporal, and striatal proton relaxation times in schizophrenic patients and normal comparison subjects. *Am. J. Psychiatry* **149,** 549–551 (1992).
5. Warntjes, J. B. Dahlqvist, O. & Lundberg, P. Novel method for rapid, simultaneous T1, T*2, and proton density quantification. *Magn. Reson. Med.* **57,** 528–537 (2007).
6. Warntjes, J. B. Leinhard, O. D., West, J. & Lundberg, P. Rapid magnetic resonance quantification on the brain: optimization for clinical usage. *Magn. Reson. Med.* **60,** 320–329 (2008).
7. Schmitt, P. *et al.* Inversion recovery TrueFISP: quantification of T1, T2, and spin density. *Magn. Reson. Med.* **51,** 661–667 (2004).
8. Ehses, P. *et al.* IR TrueFISP with a golden-ratio-based radial readout: fast quantification of T1, T2, and proton density. *Magn. Reson. Med.* **69,** 71–81 (2013).
9. Donoho, D. L. Compressed sensing. *IEEE Trans. Inform. Theory* **52,** 1289–1306 (2006).
10. Candes, E. J. & Tao, T. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory* **52,** 5406–5425 (2006).
11. Lustig, M., Donoho, D. L. & Pauly, J. M. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58,** 1182–1195 (2007).
12. Bilgic, B., Goyal, V. K. & Adalsteinsson, E. Multi-contrast reconstruction with Bayesian compressed sensing. *Magn. Reson. Med.* **66,** 1601–1615 (2011).
13. Smith, D. S. *et al.* Robustness of quantitative compressive sensing MRI: the effect of random undersampling patterns on derived parameters for DCE- and DSC-MRI. *IEEE Trans. Med. Imaging* **31,** 504–511 (2012).
14. Deshpande, V. S., Chung, Y.-C., Zhang, Q., Shea, S. M. & Li, D. Reduction of transient signal oscillations in True-FISP using a linear flip angle series magnetization preparation. *Magn. Reson. Med.* **49,** 151–157 (2003).
15. Çukur, T. Multiple repetition time balanced steady-state free precession imaging. *Magn. Reson. Med.* **62,** 193–204 (2009).
16. Nayak, K. & Lee, H. Wideband SSFP: alternating repetition time balanced steady state free precession with increased band spacing. *Magn. Reson. Med.* **58,** 931–938 (2007).
17. Lee, K., Lee, H. & Hennig, J. Use of simulated annealing for the design of multiple repetition time balanced steady-state free precession imaging. *Magn. Reson. Med.* **68,** 220–226 (2012).
18. Ernst, R. R. Magnetic resonance with stochastic excitation. *J. Magn. Reson.* **3,** 10–27 (1970).
19. Scheffler, K. & Hennig, J. Frequency resolved single-shot MR imaging using stochastic k-space trajectories. *Magn. Reson. Med.* **35,** 569–576 (1996).
20. Haldar, J. P., Hernando, D. & Liang, Z.-P. Compressed-sensing MRI with random encoding. *IEEE Trans. Med. Imaging* **30,** 893–903 (2011).
21. Doneva, M. *et al.* Compressed sensing reconstruction for magnetic resonance parameter mapping. *Magn. Reson. Med.* **64,** 1114–1120 (2010).
22. Stoecker, T., Vahedipour, K., Pracht, E., Brenner, D. & Shah, N. J. in *Proc. 19th Scientific Meeting International Society for Magnetic Resonance in Medicine* 381 (Int. Soc. Magn. Reson. Med., 2011).
23. Davenport, M. A., Wakin, M. B. & Baraniuk, R. G. *The Compressive Matched Filter* (Tech. Rep. TREE 0610, Rice University, 2006).
24. Tropp, J. A. & Gilbert, A. C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory* **53,** 4655–4666 (2007).
25. Schmitt, P. *et al.* A simple geometrical description of the TrueFISP ideal transient and steady-state signal. *Magn. Reson. Med.* **55,** 177–186 (2006).
26. Lee, J. H., Hargreaves, B. A., Hu, B. S. & Nishimura, D. G. Fast 3D imaging using variable-density spiral trajectories with applications to limb perfusion. *Magn. Reson. Med.* **50,** 1276–1285 (2003).
27. Marseille, G., De Beer, R., Fuderer, M., Mehlkopf, A. & Van Ormondt, D. Nonuniform phase-encode distributions for MRI scan time reduction. *J. Magn. Reson. B.* **111,** 70–75 (1996).
28. Tsai, C. M. & Nishimura, D. G. Reduced aliasing artifacts using variable-density K-space sampling trajectories. *Magn. Reson. Med.* **43,** 452–458 (2000).
29. Vymazal, J. *et al.* T1 and T2 in the brain of healthy subjects, patients with Parkinson disease, and patients with multiple system atrophy: relation to iron content. *Radiology* **211,** 489–495 (1999).
30. Deoni, S. C. L., Peters, T. M. & Rutt, B. K. High-resolution T1 and T2 mapping of the brain in a clinically acceptable time with DESPOT1 and DESPOT2. *Magn. Reson. Med.* **53,** 237–241 (2005).
31. Whittall, K. P. *et al. In vivo* measurement of T2 distributions and water contents in normal human brain. *Magn. Reson. Med.* **37,** 34–43 (1997).
32. Poon, C. S. & Henkelman, R. M. Practical T2 quantitation for clinical applications. *J. Magn. Reson. Imaging* **2,** 541–553 (1992).
33. Haacke, E. M., Brown, R. W., Thompson, M. R. & Venkatesan, R. *Magnetic Resonance Imaging: Physical Principles and Sequence Design* 669–675 (Wiley & Sons, 1999).
34. Hahn, E. L. Spin echoes. *Phys. Rev.* **80,** 580–594 (1950).
35. Crawley, A. P. & Henkelman, R. M. A Comparison of one-shot and recovery methods in T1 imaging. *Magn. Reson. Med.* **7,** 23–34 (1988).
36. Deoni, S. C. L., Rutt, B. K. & Peters, T. M. Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. *Magn. Reson. Med.* **49,** 515–526 (2003).
37. Scheffler, K. & Lehnhardt, S. Principles and applications of balanced SSFP techniques. *Eur. Radiol.* **13,** 2409–2418 (2003).
38. Needell, D. & Tropp, J. A. Cosamp: iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* **26,** 301–321 (2009).
39. Goldstein, T. & Osher, S. The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.* **2,** 323–343 (2009).
40. Chartrand, R. & Yin, W. in *Proc. ICASSP 2008 IEEE International Conference* 3869–3872 (IEEE, 2008).
41. Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S. & Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31,** 210–227 (2009).
42. Turk, M. & Pentland, A. Eigenfaces for recognition. *J. Cogn. Neurosci.* **3,** 71–86 (1991).
43. Griswold, M. A. *et al.* Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med.* **47,** 1202–1210 (2002).
44. Pruessmann, K. P. *et al.* SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.* **42,** 952–962 (1999).
45. Seiberlich, N., Ehses, P., Duerk, J., Gilkeson, R. & Griswold, M. A. Improved radial GRAPPA calibration for real-time free-breathing cardiac imaging. *Magn. Reson. Med.* **65,** 492–505 (2011).
46. Heidemann, R. M. *et al.* Direct parallel image reconstructions for spiral trajectories using GRAPPA. *Magn. Reson. Med.* **56,** 317–326 (2006).
47. Barger, A. V., Block, W. F., Toropov, Y., Grist, T. M. & Mistretta, C. A. Time-resolved contrast-enhanced imaging with isotropic resolution and broad coverage using an undersampled 3D projection trajectory. *Magn. Reson. Med.* **48,** 297–305 (2002).
48. Perlin, K. An image synthesizer. *Comput. Graphics* **19,** 287–296 (1985).
49. Hargreaves, B. A., Nishimura, D. G. & Conolly, S. M. Time-optimal multidimensional gradient waveform design for rapid imaging. *Magn. Reson. Med.* **51,** 81–92 (2004).
50. Fessler, J. A. & Sutton, B. P. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Trans. Signal Process.* **51,** 560–574 (2003).
51. Riffe, M. J., Blaimer, M., Barkauskas, K. J., Duerk, J. L. & Griswold, M. A. in *Proc. 15th Scientific Meeting, International Society for Magnetic Resonance in Medicine* 1879 (Int. Soc. Magn. Reson. Med, 2007).
52. Lin, L. I.-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45,** 255–268 (1989).

Supplementary Information is available in the online version of the paper.

Acknowledgements Support for this study was provided by NIH R01HL094557 and Siemens Healthcare. We also thank H. Saybasili and G. Lee for technical assistance during the implementation of these concepts; M. Lustig and W. Grissom for discussions regarding this work; and A. Exner, S. Brady-Kalnay, E. Karathanasis, E. Lavik and H. Salz for their assistance in preparing the manuscript.

Author Contributions D.M., concept development, technical implementation, data collection and analysis, manuscript development and editing; V.G., concept development, manuscript development and editing; N.S., concept development, manuscript development and editing; K.L., concept development, technical implementation, manuscript development and editing; J.L.S., concept development, manuscript development and editing; J.L.D., concept development, manuscript development and editing; M.A.G., concept development, data collection and analysis, manuscript development and editing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.G. (mark.griswold@case.edu).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Sequence design. After an initial inversion pulse, the first sequence pattern shown in Fig. 1c used a pseudorandomized series (Perlin noise[48]) of flip angle and a random repetition time between 10.5 ms and 14 ms based on a uniform random number generator. A linear ramp was added to the flip angle train since we have seen that this can increase differential sensitivity to both T_1 and T_2.

The second flip angle pattern in Fig. 1d used a series of repeating sinusoidal curves with a period of 250 repetition times and alternating maximum flip angles. In the odd periods, the flip angle (FA) is calculated as $FA_t = 10 + \sin(\frac{2\pi}{500}t) \times 50 + \text{random}(5)$, where t is from 1 to 250, and random(5) is a function to generate uniformly distributed values with a standard deviation of 5. In the even periods, we divide the previous period's flip angle by 2. A 600 ms delay was added between each of the periods to allow for both differential magnetization recovery according to T_1 and differential signal decay according to T_2. In this case, the repetition time was a Perlin noise pattern. The radio frequency phase for both of the patterns in Fig. 1 alternated between 0 and 180° on successive pulses.

The variable density spiral-out trajectory was designed to have 5.8 ms readout time in each repetition time and to have zero and first moment gradient compensation using minimum-time gradient design[49]. (The code used for this design is available at http://www-mrsrl.stanford.edu/~brian/mintgrad/). This trajectory required one interleaf to sample the inner 10×10 region, while 48 interleaves were required to fully sample the outer portions of k-space. During acquisition, the spiral trajectory rotated 7.5° from one time point to the next, so that each time point had a slightly different spatial encoding.

Dictionary design. The dictionary used in the matching algorithm was simulated using MATLAB (The MathWorks). Signal time courses with different sets of characteristic parameters (T_1, T_2 and off-resonance) were simulated. The ranges of T_1 and T_2 for the *in vivo* study were chosen according to the typical physiological limits of tissues in the brain: T_1 values were taken to be between 100 and 5,000 ms (in increments of 20 ms below a T_1 of 2,000 ms and in an increment of 300 ms above). The T_2 values included the range between 20 and 3,000 ms (with an increment of 5 ms below a T_2 of 100 ms, an increment of 10 ms between 100 ms and 200 ms, and an increment of 200 ms above a T_2 of 200 ms). Since magnetic resonance is sensitive to parts per million (p.p.m.) level deviations in the B_0 field, different off-resonance frequencies (1 Hz increment between ±40 Hz, 2 Hz between ±40 to ±80 Hz, 10 Hz between ±90 to ±250 Hz, and 20 Hz between ±270 to ±400 Hz) were simulated for each combination of T_1 and T_2 parameters to incorporate the effects of signal evolutions in different B_0 fields. A total of 563,784 dictionary entries, each with 1,000 time points, were generated in 399 s on a standard desktop computer. One dictionary entry was selected for each measured pixel location using template matching. In this case, the vector dot-product was calculated between the measured time course and all dictionary entries (appropriately normalized to each having the same sum squared magnitude) using the complex data for both. The dictionary entry with the highest dot-product was then selected as most likely to represent the true signal evolution. The proton density (M_0) of each pixel was calculated as the scaling factor between the measured signal and the simulated time course from the dictionary. For this experiment, four parameters were retrieved simultaneously from each of the 128×128 pixels using MRF. This calculation required about 3 min on a standard desktop computer.

Data acquisition. All MRI and MRF data were acquired on a 1.5 T whole body scanner (Siemens Espree, Siemens Healthcare) with a 32 channel head receiver coil (Siemens Healthcare). A square field of view of 300 mm × 300 mm was covered with a matrix of 128×128 pixels. The slice thickness was 5 mm. Images from each acquisition block were reconstructed separately using non-uniform Fourier transform (NUFFT)[50]. The resultant time series of images was used to determine the value for the parameters (T_1, T_2, M_0 and off-resonance) as described above.

In vivo experiments were performed with IRB guidelines, including written informed consent. For the fully sampled spiral acquisition shown in Supplementary Video 1, 48 repetitions were acquired, each with a different interleaf of the total acquisition. A recovery time of 5 s was used in between various acquisitions and this was taken into account in the simulated dictionary.

For the phantom study shown in Figs 2 and 4, eight cylindrical phantoms were constructed with varying concentrations of $GdCl_3$ (Aldrich) and agarose (Sigma) to yield different T_1 and T_2 values ranging from 67 to 1,700 ms and 30 to 200 ms, respectively. Standard spin echo sequences were used to quantify T_1 and T_2 separately (T_1 quantification: 13 repetition times (TRs) ranging from 50 to 5,000 ms, echo time TE = 8.5 ms, total acquisition time = 33.4 min; T_2 quantification: spin echo sequences with TEs = [15, 30, 45, 60, 90, 150, 200, 300, 400] ms, TR = 10,000 ms, total acquisition time = 3.2 h). T_1 values were calculated pixel-wise using a standard three-parameter nonlinear least squares fitting routine to solve the equation: $S(\text{TR}) = a + be^{\text{TR}/T1}$. T_2 values were determined in a pixel-wise fashion using a two-parameter nonlinear least squares fitting routine to solve the equation $S(\text{TE}) = ae^{-\text{TE}/T2}$, DESPOT1 and DESPOT2 sequences using a fully sampled spiral readout were implemented based on the acquisition values from ref. 30: DESPOT1: FA: 4° and 15°, TR: 13.6 ms, DESPOT2: FA: 15° and 55°, TR = 10.8 ms. The T_1 and T_2 values were calculated from the equations provided in ref. 30. A 20 s waiting period was used in between the different acquisitions. The initial 10 s of data acquisition was not used in order to ensure that the signal was in steady-state for each of the DESPOT acquisitions. In the following analysis of efficiency, only the pure time of data acquisition for the steady-state DESPOT images is used. For DESPOT1 this was 1.27 s and for DESPOT2 it was 2.29 s (which includes the time for the required DESPOT1 acquisition.)

Statistical analysis. Quantitative estimates of the errors and efficiencies of MRF, DESPOT1 and DESPOT2 were calculated pixel-wise using a bootstrapped Monte Carlo method[51]. Two sets of raw data were acquired for each sequence: the encoded signal and a separate acquisition that only contained noise. Fifty reconstructions were then calculated by randomly resampling the acquired noise and adding it to the raw data before reconstruction and quantification. The means and standard deviations of T_1 and T_2 along the 50 repetitions were calculated, and both were averaged within a 5 pixel × 5 pixel square region of interest for each phantom. The concordance correlation coefficients (ρ_c) were calculated using the equation[52]:

$$\rho_c = \frac{2S_{12}}{S_1^2 + S_2^2 + (\overline{Y_2} - \overline{Y_1})^2}$$

where Y_1 and Y_2 denotes the T_1 or T_2 values from two different methods, n is the number of phantoms, $\overline{Y_j} = \frac{1}{n}\sum_{i=1}^{n} Y_{ij}$, $S_j^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_{ij} - \overline{Y_j})^2$, $j = 1,2$ and $S_{12} = \frac{1}{n}\sum_{i=1}^{n}(Y_{i1} - \overline{Y_1})(Y_{i2} - \overline{Y_2})$.

The efficiency of the methods was calculated using:

$$\text{Efficiency} = \frac{T_n NR}{\sqrt{T_{\text{seq}}}},\ n = 1,2$$

where T_nNR is the T_1 or T_2 to noise ratio (defined as the T_1 or T_2 value divided by the estimated error). T_{seq} is the total acquisition time for MRF, and the relevant acquisition times for DESPOT1 and DESPOT2 (where the waiting times required for the approach to steady state and the time between each of the DESPOT1 and DESPOT2 scans to allow for complete recovery of magnetization were ignored).

©2013 Macmillan Publishers Limited. All rights reserved

ARTICLE

OPEN

doi:10.1038/nature11968

Patterns of population epigenomic diversity

Robert J. Schmitz[1,2]*, Matthew D. Schultz[2,3]*, Mark A. Urich[1,2], Joseph R. Nery[2], Mattia Pelizzola[2]†, Ondrej Libiger[4], Andrew Alix[1], Richard B. McCosh[1], Huaming Chen[2], Nicholas J. Schork[4] & Joseph R. Ecker[1,2,5]

Natural epigenetic variation provides a source for the generation of phenotypic diversity, but to understand its contribution to such diversity, its interaction with genetic variation requires further investigation. Here we report population-wide DNA sequencing of genomes, transcriptomes and methylomes of wild *Arabidopsis thaliana* accessions. Single cytosine methylation polymorphisms are not linked to genotype. However, the rate of linkage disequilibrium decay amongst differentially methylated regions targeted by RNA-directed DNA methylation is similar to the rate for single nucleotide polymorphisms. Association analyses of these RNA-directed DNA methylation regions with genetic variants identified thousands of methylation quantitative trait loci, which revealed the population estimate of genetically dependent methylation variation. Analysis of invariably methylated transposons and genes across this population indicates that loci targeted by RNA-directed DNA methylation are epigenetically activated in pollen and seeds, which facilitates proper development of these structures.

DNA methylation is a covalent base modification of plant nuclear genomes that is accurately inherited through both mitotic and meiotic[1] cell divisions. However, similarly to spontaneous mutations in DNA, errors in the maintenance of methylation states result in the accumulation of single methylation polymorphisms (SMPs) over an evolutionary timescale[2,3]. The rates of SMP formation are orders of magnitude greater than those of spontaneous mutations, which are in part, probably due to the lower fidelity of maintenance DNA methyltransferases and accompanying silencing machinery[2–5]. Epiallele formation in the absence of genetic variation can result in phenotypic variation, which is most evident in the plant kingdom, as exemplified by the *peloric* and *colorless non-ripening* variants from *Linaria vulgaris* and *Solanum lycopersicum*, respectively[6,7]. Although rates of spontaneous variation in DNA methylation and mutation can be decoupled in the laboratory[8–11], in natural settings, these two features of genomes co-evolve to create phenotypic diversity on which natural selection can act. In plant genomes, DNA methylation is present in the symmetrical CG and CHG contexts (where H = A, C or T) as well as the asymmetrical CHH context. CG gene-body methylation is a common feature of animal and plant genomes[12,13]. Regions of plant genomes that contain methylation in the CG, CHG and CHH contexts are indicative of loci that are under control of RNA-directed DNA methylation (RdDM)[14].

Similarly to the limited examples of pure epialleles (methylation variants that form independent of genetic variation), few examples of DNA methylation variants linked to genetic variants are known[15–17]. Previous studies comparing two accessions of *A. thaliana* or *Zea mays* revealed genome-wide natural variation in DNA methylation[18–21], but the dependence of these methylation variants on genetic variants at the population level remains unaddressed. To understand the types and extent of natural DNA methylation variants in *A. thaliana*, epigenomes for genotypically distinct, wild accessions, isolated from throughout the Northern Hemisphere, were determined using MethylC-sequencing[18], (152 methylomes, Supplementary Table 1), RNA-sequencing (144 transcriptomes, Supplementary Table 2) and genomic DNA-sequencing (217 genomes, Supplementary Table 3)[18]. Integration of genomic and epigenomic data allowed investigation into variable methylation states of both CG gene-body methylation and loci targeted by RdDM along with their interactions with genetic variants at the population level.

Population-wide patterns of SMPs

Recent reports of SMPs in a population of essentially isogenic plants indicated that they are major contributors to epigenomic variation[2,3]. Therefore, we assessed SMP diversity to understand their frequency and patterns throughout a population of genetically distinct accessions. A median of 390,255 SMPs ranging from 92,646 to 527,393 (Supplementary Table 4) were found in the sequenced accessions when compared to the Col-0 reference methylome. On average, CG-, CHG- and CHH-SMPs accounted for 23%, 13% and 64% of all SMPs, respectively. These newly identified SMPs were used to construct an epigenome-based phylogeny and then were compared to a genome-based (single nucleotide polymorphism (SNP)) phylogeny (Supplementary Figs 1–4). A high correlation in the tree structures was specifically observed between CG-SMPs and SNPs as compared to CHG-SMPs or CHH-SMPs and SNPs (Supplementary Table 5).

To determine patterns of SMP diversity, chromosome-wide conservation of methylation states at each SMP was examined by computing a conservation score (Fig. 1a and Supplementary Fig. 5). The methylation state of SMPs in the CG and CHG contexts is biased towards the methylated form at the pericentromere and biased towards the unmethylated form in gene-rich regions (Fig. 1a and Supplementary Fig. 5). Next, the distribution of conservation scores across different features and methylation contexts were plotted genome-wide (Fig. 1b–d). Like the pericentromeric regions, CG- and CHG-SMPs in transposable elements tend to be faithfully methylated throughout this population, whereas CHH-SMPs are largely unmethylated. Unlike CHG and CHH-SMPs, CG-SMPs have a significantly larger amount of methylation at single-copy genes (Fig. 1b–d). Because CG gene-body methylation is associated with moderately expressed genes[19], we postulated that these genes are more active because of the lack of other genes redundant in function. We tested this hypothesis by examining RNA-seq data for 144 of these accessions

[1]Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. [2]Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. [3]Bioinformatics Program, University of California at San Diego, La Jolla, California 92093, USA. [4]The Scripps Translational Science Institute and the Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037, USA. [5]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.
†Present address: Center for Genomic Science of IIT@SEMM, Istituto Italiano di Tecnologia (ITT), Via Adamello 16, 20139 Milan, Italy.
*These authors contributed equally to this work.

©2013 Macmillan Publishers Limited. All rights reserved

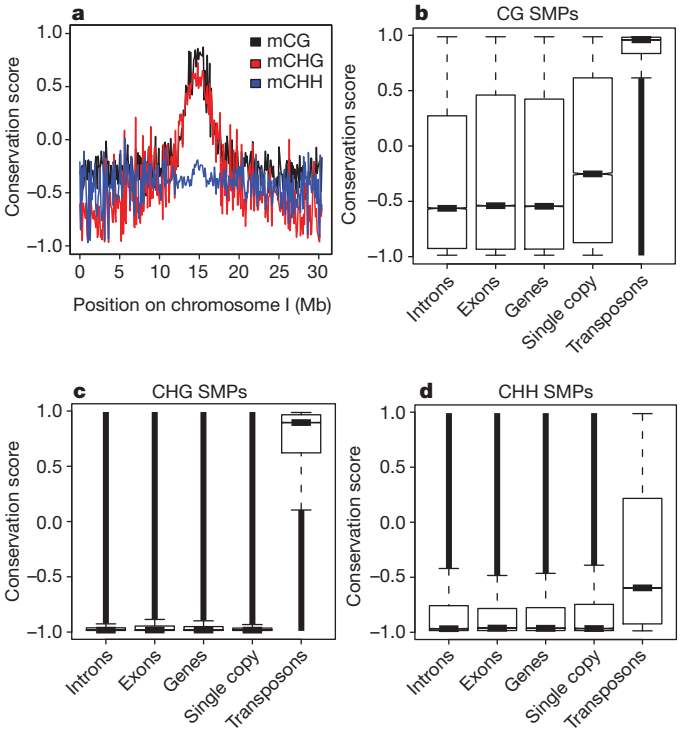


Figure 1 | Population-wide analyses of SMPs. **a**, A plot of the genome-wide distribution of methylation conservation across chromosome I. Mb, megabases. **b**–**d**, The distributions of SMP conservation scores across various genomic features. Notches in the boxplots represent bootstrap estimated 95% confidence intervals around the medians. mCG, methylated cytosine in CG context; mCHG, methylated cytosine in CHG context; and mCHH, methylated cytosine in CHH context.

at these loci, which revealed the fraction of transcripts where expression was detected—that is, FPKMs > 0, where FPKM indicates fragments per kilobase of exon per million fragments mapped—was higher in single-copy genes than non-single-copy genes (85% versus 71.8%). Moreover, the median expression level of single-copy genes was also significantly greater (361,814.50 FPKMs of single-copy genes versus 56,107.85 FPKMs of non-single-copy genes), supporting the finding that single-copy genes across the population are more transcriptionally active.

Population-wide variation of DMRs

Spontaneous formation of SMPs represents one form of natural epigenetic variation, but variation also exists in the form of differentially methylated regions (DMRs)[2,3]. Therefore, we scanned this population for DMRs in the CG context (CG-DMRs) typically found in gene-bodies or in the CG, CHG and CHH contexts (C-DMRs) typical of regions targeted by RdDM. Because CG-DMRs and C-DMRs are not mutually exclusive, only CG-DMRs that did not overlap with a C-DMR were defined as CG-DMRs. Hierarchical clustering of accessions based on weighted methylation levels[20] (Supplementary Information)—referred to as methylation levels throughout the rest of the paper—of CG-DMRs or C-DMRs revealed patterns across the population that were coincident with certain genomic features (Fig. 2a and b). For example, CG-DMRs are enriched in gene bodies and are present in both unmethylated and methylated states equally throughout the population (Fig. 2a), whereas C-DMRs occur in both gene bodies and transposons (Fig. 2b). Additionally, the C-DMRs in genes are largely unmethylated, which contrasts with the heavy methylation levels that occur in transposons (Fig. 2b). In total, 40,269 CG-DMRs (Supplementary Table 6), with an average size of 321 base pairs (bp; Supplementary Fig. 6) were identified across the population that were enriched in gene bodies and depleted in transposons (Fig. 2a, Supplementary Fig. 7 and Supplementary Table 7). We also identified 13,485 C-DMRs (Supplementary Table 8), with an average size of 221 bp (Supplementary Fig. 6), that showed enrichment in transposons and depletion in genes (Fig. 2b, Supplementary Fig. 7 and Supplementary Table 7).

This distribution of both CG- and C-DMRs reflects the distribution of genes and transposons along each chromosome and the type of DNA methylation primarily associated with these features, namely CG gene-body methylation versus RdDM. Furthermore, the distribution of methylation levels of CG-DMRs is skewed towards lower levels when the CG-DMR overlaps a gene and towards higher levels when it overlaps a transposon (Fig. 2c, d). The distribution of methylation levels in CG-DMRs resembles the patterns of CG-SMPs for genes versus transposons, as the transposon sequences often contained highly methylated sites or DMRs when compared to genes, supporting the observation that these regions are faithfully repressed by methylation across the population. A comparison of the distribution of methylation levels of the C-DMRs revealed that genes are infrequently methylated at high levels in the population when compared to C-DMRs overlapping transposons (Fig. 2c, d). In this regard, C-DMRs overlapping genes are rare variants in the population, whereas most transposon sequences are almost invariably methylated. Clustering these accessions based on their methylation levels of C-DMRs revealed that accessions that are geographically separated are less likely to cluster together, indicating the potential for underlying genetic structure (Fig. 2e, f). Alternatively, these results could also be obtained for methylation variants that are not dependent on genetic variants if they are stable. Therefore, the observation that accessions cluster based on their geographical distribution is probably due to a combination of both genetic structure and stable methylation variants.

For a subset of accessions examined, methylation data were produced for two tissue types: leaf and mixed-stage inflorescence. Regardless of the tissue used for methylome analysis, when hierarchical clustering was performed using methylation levels of either CG-DMRs (Fig. 2g) or C-DMRs (Fig. 2h), these accessions grouped by their genotype not their tissue type. When the same analysis was applied to RNA-seq data from the same tissues of six accessions, samples clustered on the basis of their tissue type not their genotype (Fig. 2i). Collectively, these data indicate that DNA methylation is less dynamic than gene expression patterns in plants and only plays a role during specific stages of development or cell types[1,21,22]. Although DNA methylation is more static than transcription, it varies appreciably over an evolutionary timescale, significantly affecting the transcriptional output of specific genes (Fig. 2j, k). Using CG-DMRs that overlap with genes, a positive correlation (Spearman correlation; $P < 2.2 \times 10^{-16}$) between their methylation levels and gene expression levels were found (Fig. 2j), whereas the opposite was true for C-DMRs that overlapped genes, supporting a role for RdDM in transcriptionally silencing these loci (Spearman correlation; $P < 2.2 \times 10^{-16}$: Fig. 2k, Supplementary Figs 8 and 9, and Supplementary Information). Although the role of CG gene-body methylation is still unclear, these data indicate that CG-DMRs that are heavily methylated are associated with higher gene expression levels and can possibly give rise to transcriptional variation.

Linking genetic and methylation variants

Genome sequencing was performed for 217 individuals of which 152 had a matching sequenced DNA methylome. We used the SHORE analysis pipeline[23] to identify SNPs between each accession and the Col-0 genome (Supplementary Information). The identification of SMPs and SNPs that were variable between at least two accessions was used to determine the population-level frequency of these variants, which revealed approximately 70% of CG-SMPs and 41% of SNPs are present at <1% allele frequency (Supplementary Table 9). These results indicate that a large fraction of SMPs and SNPs are

©2013 Macmillan Publishers Limited. All rights reserved

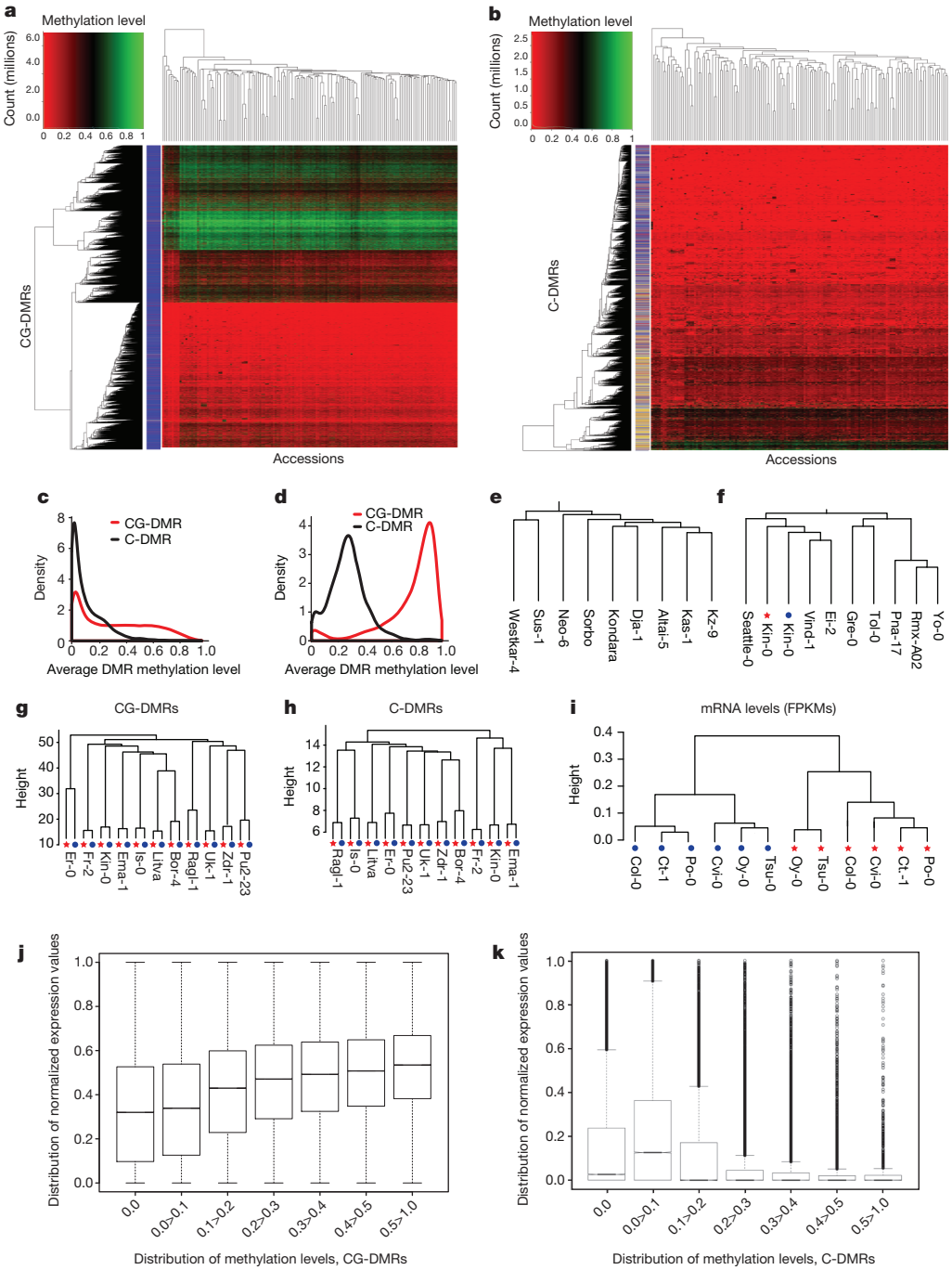


Figure 2 | Population-wide analyses of DMRs. **a**, **b**, Heatmaps representing the methylation levels across CG-DMRs (**a**) and C-DMRs (**b**) with coverage in all 152 accessions. The coloured row labels on the left of the heatmap indicate what feature a DMR falls in (blue, gene; gold, transposon; red, gene with a transposon inserted in an intron; grey, no feature). Rows indicate genomic locus of DMR and columns indicate accessions. **c**, **d**, The density (*y* axis; arbitrary units) and average methylation levels (*x* axis) of CG-DMRs and C-DMRs in genes (**c**) and transposons (**d**). **e**, **f**, Asian (**e**) and North American (**f**) methylome profiles reflected geographical distribution. **g**–**i**, Dendrograms from hierarchical clustering of CG-DMRs (**g**), C-DMRs (**h**) and mRNA levels (**i**) from accessions that had samples from two different tissues (*y* axis, height in arbitrary units). Red stars and blue circles indicate leaf and mixed stage inflorescence samples, respectively. **j**, Boxplot representation of transcriptional variation reveals a positive association with higher levels of methylation levels of CG-DMRs. **k**, Increasing methylation levels of C-DMRs are negatively associated with gene expression.

rare variants similar to the results observed for C-DMRs and further indicate that the high epimutation rate for SMPs results in greater numbers of rare alleles. Therefore, even though the spontaneous epimutation rate is at least four orders of magnitude greater than SNPs, the reversible nature of certain SMPs governs their accumulation within populations[2,3,5].

Our analysis of gene families that contained the highest number of major effect mutations (see Supplementary Information: NBS-LRR, for defence response; F-box, for protein degradation; and MADS-box transcription factor, for development) is consistent with previous studies[24,25], and these gene families also contained the highest frequency of C-DMRs (Fig. 3a). Furthermore, gene ontology analysis for genes overlapping with C-DMRs identified terms enriched in protein degradation and immune response functions, indicating that these three gene families are equally prone to hypervariable genetic and epigenetic states (that is, they are hypermutable; Supplementary Table 10). Although the frequency of major effect mutations and C-DMRs was similar for these hypermutable families, the remaining

©2013 Macmillan Publishers Limited. All rights reserved

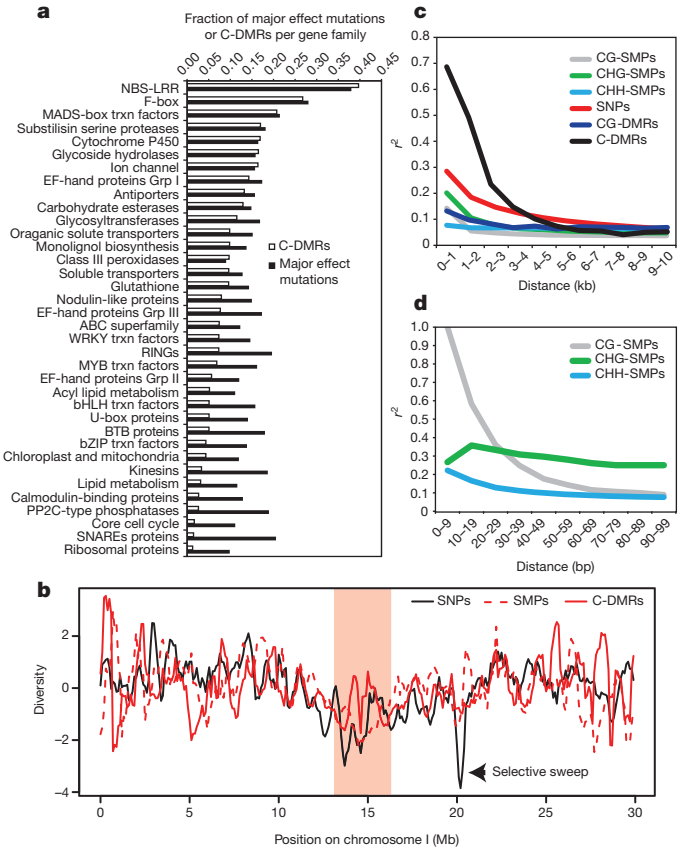


Figure 3 | Population-wide analyses of natural genetic and epigenetic variation. **a**, A comparison of genetic and epigenetic variation across gene families that have the highest incidence of C-DMRs. The *y* axis indicates the gene family and the *x* axis represents the relative fraction of major effect mutations or fraction of C-DMRs. Fractions were used to adjust for the size of the gene families. trxn, transcription. **b**, A plot of SNP, SMP and C-DMR diversity across chromosome I. The shaded pink region indicates the position of the pericentromere. **c**, **d**, Linkage disequilibrium/positional association decay plots for genetic and epigenetic variants.

gene families tested revealed no such co-occurrence of genetic and methylation variation as the frequency of C-DMRs approached zero, whereas the frequency of major effect mutations reached a background rate (Fig. 3a). Apart from the hypermutable families discussed above, there is little relationship between major effect mutations and frequency of C-DMRs. Furthermore, there is no correlation between methylation level and mutation rate in genes containing C-DMRs (Supplementary Table 11). Therefore, the majority of genes targeted by RdDM are functional, and silencing by this pathway may limit their expression to specific stages of development (similar to observations made for transposons[26]) and/or limit their expression until released from silencing by bacterial infection[27], possibly explaining the high frequency of C-DMRs in members of the NBS-LRR family.

To determine the extent to which variation in both DNA methylation and genotype are linked, diversity estimates were calculated for SNPs, all forms of SMPs and C-DMRs (Fig. 3b and Supplementary Fig. 10). A known selective sweep on chromosome I (ref. 26) was identified (Fig. 3b). However, no corresponding depletion was observed for either CG-SMPs or C-DMRs. At this resolution, no correlation between genotype and epigenotype was detected (Supplementary Table 12). Therefore, to understand the relationship and possible dependence of methylation variants on genotype, a higher-resolution positional association and linkage disequilibrium decay analysis was performed using SNPs, CG-SMPs, CHG-SMPs, CHH-SMPs, CG-DMRs and C-DMRs (Fig. 3c, d). Similar to past reports for SNPs, linkage disequilibrium decays within 10 kilobases (kb), reaching 50% of its initial value at ~2 kb (refs 25, 28; Fig. 3c). This value is similar to the rate of decay for the association amongst C-DMRs (~10 kb), which reaches 50% of its initial value at ~1 kb (Fig. 3c). Surprisingly, the rate of decay for association amongst methylation variants such as CG-SMPs and CG-DMRs occurs rapidly, within 100 bp, which is especially true for genes when compared to transposons (Fig. 3d and Supplementary Figs 11 and 12). Collectively, these data indicate that SMPs and CG-DMRs are truly epigenetic in nature, as they occur largely independent of genetic variation. In contrast, although spontaneous C-DMR formation can occur independent of genetic variation[2,3], the linkage disequilibrium and association decay analysis revealed that the presence of C-DMRs may be due, in part, to local genetic variants.

Association-mapping methylation variants

Although there are many mechanisms that can give rise to DNA methylation variation[2,3,15,29], the extent to which each plays a role in the formation of the observable methylation variation is unknown. We noted that some sites of known transposition events possessed C-DMRs and posited that these structural variants could be responsible for these differences (Supplementary Fig. 13). To experimentally determine the proportion of C-DMRs with a local structural variant, regions surrounding 92 C-DMRs were PCR amplified and sequenced. Most of these C-DMRs failed to overlap with structural variants; however, structural variations were detected at ~17% (16/92) of the C-DMRs assayed (Fig. 4a, Supplementary Table 13). To better inspect any direct relationship between genetic variants and C-DMRs and to identify potential methylation quantitative trait loci (mQTL)[30], we used a genome-wide association technique, EMMAX, as this methodology was successfully used in another similarly sized *Arabidopsis* population[28,31] (Supplementary Information). Furthermore, we employed two different methodologies to control for false discoveries and found them highly concordant (Supplementary Information). To minimize the number of false positives, we used SNPs that were significant in both methodologies. Application of EMMAX to the 152 accessions

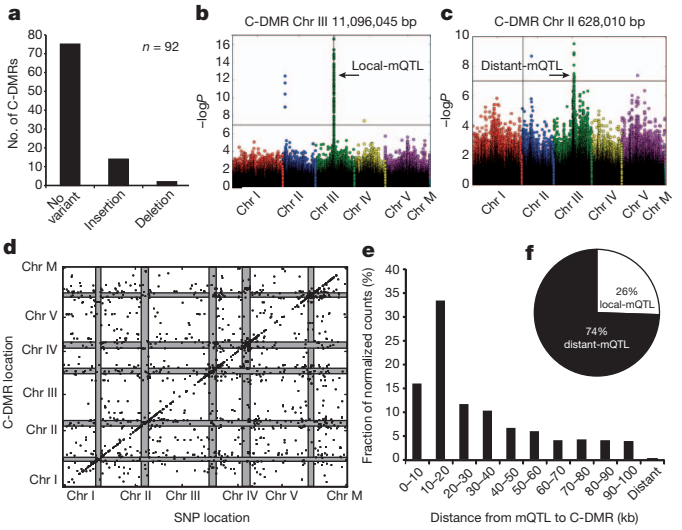


Figure 4 | Association of natural genetic variants and methylation variants. **a**, A summary of the type and number of variants (non-SNPs and small indels) discovered at 92 C-DMRs. **b**, **c**, Manhattan plots with examples of local (**b**) and distant (**c**) mQTL. Chr M denotes the mitochondrial genome. **d**, Distribution of significant mQTL and the C-DMRs with which they associate. Each point represents a significant association between a C-DMR and a block of SNPs. The *x* axis denotes the genomic location of the SNP block, and the *y* axis indicates the position of the C-DMR. The pericentromeres on each chromosome are shown as grey bars. **e**, The distribution of distances of mQTL from their C-DMRs normalized for the base space covered by each range of distances. **f**, The ratio of distant mQTL to local mQTL.

©2013 Macmillan Publishers Limited. All rights reserved

with SNP and C-DMR data uncovered C-DMRs that associated with local (Fig. 4b) and distant genetic variants (Fig. 4c) and identified the well-characterized *PAI* epialleles (Supplementary Fig. 14)[15]. In total, 2,739 significant mQTL were associated with 1,045 of the 3,023 tested C-DMRs (~35%; Supplementary Figs 15–24).

Of the tested C-DMRs, 377 (~12%) overlap with a genomic locus with which they associate, which is a similar proportion to the number of experimentally determined local variants. We grouped significant mQTL into blocks and plotted the position of these blocks and the corresponding C-DMR in Fig. 4d (Supplementary Information). An enrichment of local mQTL is visible in particular at the pericentromeric regions (Fig. 4d). When corrected for the genome space in which local events can occur, local mQTL account for a larger fraction of the overall results, although the raw number of distant mQTL exceeds the number of local mQTL (Fig. 4f). Furthermore, 61.3% of the local mQTL occur within 30 kb of the C-DMR (Fig. 4e). These association-mapping results also indicated that there were more than twice as many mQTL as C-DMRs. To address whether or not many of the C-DMRs are being controlled in a polygenic manner, we applied the tool MLMM[32] to the 1,045 C-DMRs with at least one mQTL. Roughly half of the significant C-DMRs reported as polygenic by EMMAX were also reported as polygenic by MLMM (Supplementary Fig. 25). Given these results, there are polygenic C-DMRs, although it remains to be determined what types of mechanisms lead to the methylation variation of these C-DMRs. Lastly, applying EMMAX to CG-DMRs resulted in a much lower detection rate of mQTL (Supplementary Table 14 and Supplementary Fig. 26). Together, the above data demonstrate that a considerable fraction of C-DMRs and to a much lesser extent CG-DMRs exist as a result of genetic variation.

All C-DMRs randomly selected for Fig. 4a are rare in the population and had been filtered out before association mapping. Consequently, to determine potential causal variants that are associated with the methylation variants, we PCR amplified 96 C-DMRs associated with a local mQTL. Of these tested loci, 86 successfully amplified and revealed 16 structural variants (Supplementary Table 15), which are similar to the results from the randomly selected C-DMRs (16/92 versus 16/86). As an alternative to structural variation, distant mQTL may result from SNPs, as reported for the *VIM1* variant in the Bor-4 accession[33]. Analysis of components with known involvements in DNA methylation within these distant mQTL regions (Supplementary Table 16) revealed *VIM3* and *AGO2* as possible causal loci. Potential causal variants for the remaining local and distant mQTL are likely to involve a combination of either SNPs or structural variations that will undoubtedly be uncovered with future whole-genome assemblies.

RdDM targets are activated in pollen

The mQTL that we have identified revealed that there is an association between some genetic variants and DNA methylation variants, especially for C-DMRs. It is well established that other genetic features, such as repeats, are important for guiding RdDM to target loci. For example, the intergenic sub-telomeric repeats 3′ to the *MEDEA* locus and the repeated SINE elements and tandem repeats around the transcription start site of *FWA* are key regulatory sequences for controlling gene expression of these loci[34,35]. Although these loci are under transcriptional control by genetic elements, these specific elements are present and invariably methylated in every accession examined. Therefore, to understand the role of regions of the epigenome that are less prone to natural epigenetic variation, we searched for loci that contained methylated alleles (methylation level ≥10%) in >90% of the accessions and identified 283 genes and 255 transposons. The expression of these loci was specifically activated in pollen (Fig. 5a and b). A previous study demonstrated that transposons are activated in the pollen vegetative nucleus, providing a substrate to generate mobile small RNAs, which can be transmitted to the sperm cells (germ line)[26].

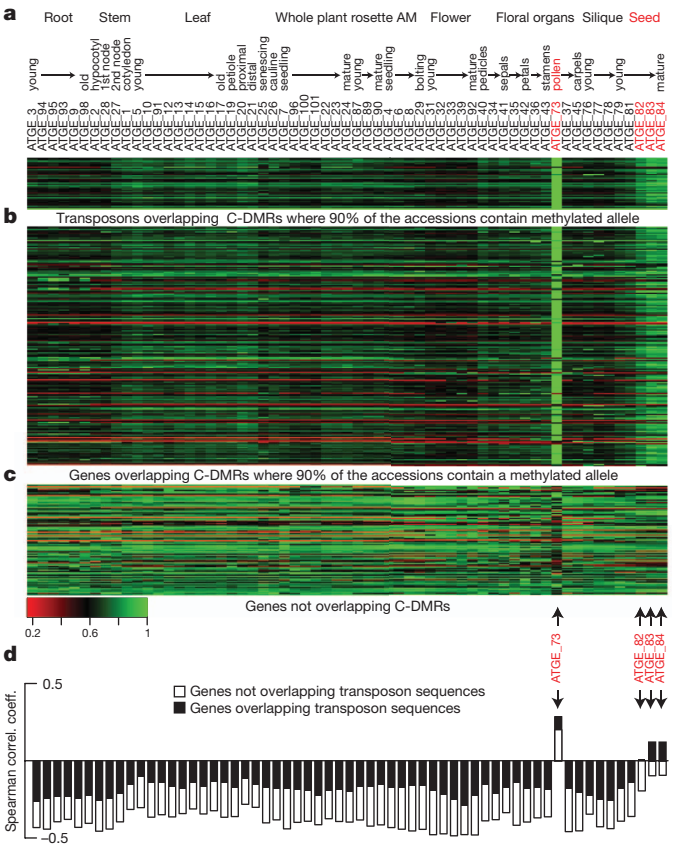


Figure 5 | Epigenetic reprogramming of genes targeted by the RdDM pathway. **a–c**, A heatmap representation of mRNA expression levels throughout a developmental time course[37] for transposons (**a**) and genes (**b**) that overlap with C-DMRs where >90% of the alleles are methylated across the population, and genes not overlapping with C-DMRs (**c**). Each row represents a locus with mRNA expression values. Each column represents a different developmental stage. AM, apical meristem. mRNA expression values on colour scale range from low (red) to medium (black) to high (green) and are represented as the average expression level across replicates normalized by the maximum average expression level for each probe. **d**, Correlation test between the nonCG methylation levels and microarray gene expression values for genes targeted by RdDM in Col-0. Open rectangles are genes that do not overlap transposon sequences; filled black rectangles represent genes overlapping transposon sequences.

This mechanism is not restricted to transposons, as we found that protein-coding genes that are under control of RdDM and invariably methylated across this population are also activated in pollen (Fig. 5b). This activation is not a general feature of pollen, as a control set of genes that are not targeted by RdDM are not activated in pollen (Fig. 5c). A closer examination of these invariably methylated genes with gene ontology revealed a significant enrichment for two major categories, cell wall biology and translation (Supplementary Table 17), both related to major functions of pollen development.

Although these invariably methylated loci are under similar epigenetic control as transposons (Fig. 5a, b), it is likely that all RdDM-targeted loci are under the control of this mechanism regardless of their variability within this population. In fact, Col-0 genes targeted by RdDM and their corresponding expression levels are positively correlated (Spearman correlation; $P = 5.81 \times 10^{-27}$) in pollen and seed development (Fig. 5d), whereas all 55 other tissues tested revealed either a negative correlation or no correlation (Fig. 5d, Supplementary Table 18). Furthermore, categories of genes with positive correlations are stronger for loci that overlap transposon sequences (Fig. 5d). These data indicate that these loci have probably come under control of sequences that are evolutionarily silenced, which acts to restrict their

©2013 Macmillan Publishers Limited. All rights reserved

expression to these specific stages of development (Fig. 5d: see Supplementary Information for an expanded discussion of this topic).

Concluding remarks

Natural epigenomic variation is widespread within *A. thaliana*, and the population-based epigenomics presented here has uncovered features of the DNA methylome that are not linked to underlying genetic variation, such as all forms of SMPs and CG-DMRs. However, C-DMRs have positional association decay patterns similar to linkage disequilibrium decay patterns for SNPs and in some cases are associated with genetic variants, but the majority of C-DMRs were not tested by association mapping due to low allele frequencies and could result from rare sequence variants. Our combined analyses of genetic and methylation variation did not uncover a correlation between major effect mutations and genes silenced by RdDM, suggesting that this pathway may target these genes for another purpose. This purpose could be to restrict expression from vegetative tissues similarly to transposons. Another possible purpose of being targeted by RdDM could be to coordinate expression specifically in pollen and in seed to ensure proper gametophytic and embryonic development. Animals also use small RNA-directed DNA methylation and heterochromatin formation mechanisms to maintain the epigenome of the germ line through the use of Piwi-interacting RNAs[36]. In both plants and animals these small RNAs are derived from the genome of companion cells, which are terminal in nature and can afford widespread reactivation of transposon and repeat sequences as they are not passed on to the next generation. Our study provides evidence that RdDM-targeted genes may have co-opted this transposon silencing mechanism to maintain their silenced state in vegetative tissues and transgenerationally, as well as to ensure proper expression important for pollen, seed and germ line development.

METHODS SUMMARY

MethylC-seq library construction. Genomic DNA was sonicated to ~100 bp using the Covaris S2 system and then purified using Qiagen DNeasy MinElute columns (Qiagen). Each sequencing library was constructed similarly to genomic DNA libraries except the ligation was performed with methylated adapters provided by Illumina. Ligation products were purified with AMPure XP beads (Beckman). Ligated DNA was bisulphite treated using the MethylCode Kit (Invitrogen) and then PCR amplified using Pfu Cx Turbo (Agilent).

Full Methods and any associated references are available in the online version of the paper.

Received 5 July 2012; accepted 30 January 2013.
Published online 6 March 2013.

1. Calarco, J. P. *et al.* Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* **151,** 194–205 (2012).
2. Becker, C. *et al.* Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* **480,** 245–249 (2011).
3. Schmitz, R. J. *et al.* Transgenerational epigenetic instability is a source of novel methylation variants. *Science* **334,** 369–373 (2011).
4. Genereux, D. P., Miner, B. E., Bergstrom, C. T. & Laird, C. D. A population-epigenetic model to infer site-specific methylation rates from double-stranded DNA methylation patterns. *Proc. Natl Acad. Sci. USA* **102,** 5802–5807 (2005).
5. Ossowski, S. *et al.* The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327,** 92–94 (2010).
6. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401,** 157–161 (1999).
7. Manning, K. *et al.* A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nature Genet.* **38,** 948–952 (2006).
8. Johannes, F. *et al.* Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* **5,** e1000530 (2009).
9. Mirouze, M. *et al.* Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* **461,** 427–430 (2009).
10. Reinders, J. *et al.* Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev.* **23,** 939–950 (2009).
11. Teixeira, F. K. *et al.* A role for RNAi in the selective correction of DNA methylation defects. *Science* **323,** 1600–1604 (2009).
12. Feng, S. *et al.* Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl Acad. Sci. USA* **107,** 8689–8694 (2010).
13. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328,** 916–919 (2010).
14. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Rev. Genet.* **11,** 204–220 (2010).
15. Bender, J. & Fink, G. R. Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*. *Cell* **83,** 725–734 (1995).
16. Woo, H. R. & Richards, E. J. Natural variation in DNA methylation in ribosomal RNA genes of *Arabidopsis thaliana*. *BMC Plant Biol.* **8,** 92 (2008).
17. Martin, A. *et al.* A transposon-induced epigenetic change leads to sex determination in melon. *Nature* **461,** 1135–1138 (2009).
18. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133,** 523–536 (2008).
19. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genet.* **39,** 61–69 (2006).
20. Schultz, M. D., Schmitz, R. J. & Ecker, J. R. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet.* **28,** 583–585 (2012).
21. Ibarra, C. A. *et al.* Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337,** 1360–1364 (2012).
22. Gehring, M., Bubb, K. L. & Henikoff, S. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* **324,** 1447–1451 (2009).
23. Ossowski, S. *et al.* Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **18,** 2024–2033 (2008).
24. Cao, J. *et al.* Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genet.* **43,** 956–963 (2011).
25. Clark, R. M. *et al.* Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317,** 338–342 (2007).
26. Slotkin, R. K. *et al.* Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136,** 461–472 (2009).
27. Dowen, R. H. *et al.* Widespread dynamic DNA methylation in response to biotic stress. *Proc. Natl Acad. Sci. USA* **109,** E2183–E2191 (2012).
28. Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465,** 627–631 (2010).
29. Ahmed, I., Sarazin, A., Bowler, C., Colot, V. & Quesneville, H. Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in *Arabidopsis*. *Nucleic Acids Res.* **39,** 6919–6931 (2011).
30. Johannes, F., Colot, V. & Jansen, R. C. Epigenome dynamics: a quantitative genetics perspective. *Nature Rev. Genet.* **9,** 883–890 (2008).
31. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature Genet.* **42,** 348–354 (2010).
32. Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genet.* **44,** 825–830 (2012).
33. Woo, H. R., Pontes, O., Pikaard, C. S. & Richards, E. J. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev.* **21,** 267–277 (2007).
34. Kinoshita, T. *et al.* One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303,** 521–523 (2004).
35. Xiao, W. *et al.* Imprinting of the *MEA* Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev. Cell* **5,** 891–901 (2003).
36. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313,** 320–324 (2006).
37. Schmid, M. *et al.* A gene expression map of *Arabidopsis thaliana* development. *Nature Genet.* **37,** 501–506 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Luo, R. Dowen and N. Krogan for critical reading of this manuscript; B. Coullahan for technical assistance with SOLiD RNA-seq; O. Loudet for seeds; and S. Ossowski, K. Schneeberger and D. Weigel for assistance in establishing a variant calling pipeline. R.J.S. was supported by NIH postdoctoral fellowships F32HG004830 and K99GM100000. M.D.S. was supported by an NSF IGERT training grant (DGE-0504645). O.L. and N.J.S. were supported by NIH/NCRR grant number UL1 RR025774. This work was supported by the NSF (MCB-0929402 and MCB-1122246), the Howard Hughes Medical Institute and the Gordon and Betty Moore Foundation (grant GBMF3034 to J.R.E.). J.R.E. is a HHMI-GBMF Investigator.

Author Contributions R.J.S., M.D.S. and J.R.E. conceived and designed the study. R.J.S., M.A.U., J.R.N., A.A. and R.B.M. performed experiments. R.J.S., M.D.S., M.P., O.L., H.C. and N.J.S. performed data analysis. R.J.S., M.D.S. and J.R.E. wrote the paper.

Author Information Genome sequence data can be downloaded from NCBI SRA (SRA012474). Processed datasets can downloaded from NCBI GEO under accession GSE43857 for MethylC-seq data and accession GSE43858 for RNA-seq data and can be viewed at http://neomorph.salk.edu/1001_epigenomes.html and http://signal.salk.edu/atg1001/index.php. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of the paper is freely available to all readers. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.R.E. (ecker@salk.edu).

This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-sa/3.0

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Plant material. Leaf and mixed stage inflorescence tissue were flash frozen in liquid nitrogen, and then the tissue was ground to a fine powder with a mortar and pestle. Leaf tissue was used for genomic and RNA-seq, and the tissues used for each MethylC-seq experiment are listed in Supplementary Table 1. DNA was isolated using a Plant DNeasy kit (Qiagen) following the manufacturer's recommendations. RNA was isolated using the Plant RNeasy kit (Qiagen) following the manufacturer's instructions.

Genomic DNA sequencing library construction. Approximately two micrograms of genomic DNA was sonicated to ~250 bp using the Covaris S2 system using the following parameters: cycle number, 2; duty cycle, 10%; intensity, 4; cycles per burst, 200; and time, 40 s. Sonicated DNA was purified with a PCR Purification MinElute column according to the manufacturer's instructions (Qiagen). Purified DNA was end repaired at room temperature for 45 min using the End-It Repair kit (Epicentre) and purified with a MinElute column (Qiagen). Purified samples were then A-tailed with dATP and Klenow 3′–5′ exo minus (New England Biolabs) for 30 min at 37 °C and then purified with a MinElute column (Qiagen). Purified DNA was then used for an overnight ligation to TruSeq barcoded adapters (Illumina) with T4 DNA ligase at 16 °C (New England Biolabs). Ligated fragments were purified twice using Ampure XP purification beads (Beckman) at 1.3× ratio of beads to sample and then PCR amplified for 15 cycles using Phusion High Fidelity DNA Polymerase (New England Bioloabs).

MethylC-seq library construction. Approximately one to three micrograms of genomic DNA was sonicated to ~100 bp using the Covaris S2 system using the following parameters: cycle number, 6; duty cycle, 20%; intensity, 5; cycles per burst, 200; and time, 60 s. Sonicated DNA was purified using DNeasy MinElute columns (Qiagen). Each sequencing library was constructed similarly to genomic DNA libraries except the ligation was performed with methylated adapters provided by Illumina. Ligation products were purified with AMPure XP beads (Beckman) at a ratio of 1.8 of beads to sample. Up to 450 ng of ligated DNA was bisulphite treated using the MethylCode kit (Invitrogen) following the manufacturer's guidelines and then PCR amplified using Pfu Cx Turbo (Agilent) using the following PCR conditions (2 min at 95 °C, 4 cycles of 15 s at 98 °C, 30 s at 60 °C, 4 min at 72 °C and 10 min at 72 °C).

RNA-seq library construction. RNA-seq libraries were prepared according to described methods[38] except for data collected for Fig. 2i. These libraries were prepared using a TruSeq RNA Sample Kit v2 (Illumina).

Sequencing. Paired-end genomic DNA and single-end MethylC-seq libraries were sequenced using the Illumina GAIIx as per manufacturer's instructions. Sequencing of genomic DNA and MethylC-seq libraries was performed up to 101 and 85 cycles, respectively. Image analysis and base calling were performed with the standard Illumina pipeline. Sequencing of RNA-seq libraries was performed on the SOLiD4 platform (Life Technologies) for 50 bp according to the manufacturer's instructions.

Variant identification. The SHORE package was used to call variants for all of our accessions[23]. The following is a list of each submodule and arguments that we ran for the strains: shore import '–v' Fastq; '–e' Shore; '–a' genomic; '–x' forward reads; '–y' reverse reads; '–o' output directory; '–n' 200, shore mapflowcell '–l' TAIR10 Reference; '–f' output directory; '–v' bwa; '–n' 5%; '–g' 3; '–c' 7; '–b' 500000, shore correct4pe '–l' input directory; '–x' 250; '–e' 1001, shore merge '–P' input directory; '–d' output directory, shore consensus '–n' accession_name; '–f' TAIR10 Reference; '–o' output directory; '–i' input directory; '–g' 4; '–q' 7; '–a'; *Arabidopsis* default scoring matrix '–b' 0.51; '–v'; '–r'. Any variant with a quality score of 25 or above was deemed significant. These variants were then substituted into the TAIR10 reference genome to create sample specific references (also referred to as SNP-substituted references) for the mapping of other data sets. In the case of the MethylC-seq mapping, we were able to map, on average, an additional 943,182 reads and allowed us to call an additional 225,894 methylated cytosines (Supplementary Table 19).

MethylC-seq sequencing analysis. Fastq files were aligned to SNP-substituted reference genomes for each accession using Bowtie[39], and custom algorithms were used for identification of mC sites as described previously[40].

RNA-seq data analysis. Bioscope version 1.3 was used to align .csfasta and .qual files to SNP-substituted reference genomes for each accession using default parameters; this allows up to 10 locations per sequenced read. Cufflinks version 1.1 was used to quantify gene expression values using the following parameters: '–F' 0; '–b'; '–N'; '–library-type' fr-secondstrand; '–G' TAIR10.gtf.

Identification of SMPs. We identified SMPs by looking for sites that either were called methylated by our pipeline, or were covered by at least five reads, which we defined as an unmethylated site. Any other site was listed as missing. A SMP was defined as any site with an accession that had a methylation state different between at least two accessions but contained the same sequence as the Col-0 reference genome.

Dendrogram construction. Throughout this work, we present various clustering results of SMPs, SNPs and DMRs. In the cases where these dendrograms are presented with a heatmap, we used the R function heatmap.2 in the gplots package with the default clustering parameters to produce the figure. The dendrograms that lack heatmaps were produced by first generating a distance matrix with R's dist function and passing this matrix to the hclust function, both with their default parameters.

Clustering comparison. To compare the results of the clustering of SMPs and SNPs, we generated distance matrices using R's dist function with the methylation statuses of SMPs as well as the alleles of the SNPs and then compared the Spearman correlation coefficients between the SNP distance matrix and each of the SMP distance matrices (Supplementary Table 5).

Identification of DMRs. All classes of DMRs were identified as previously reported[3]. CG-DMRs and C-DMRs are not mutually exclusive because C-DMRs are a subset of CG-DMRs. Consequently, for any CG-DMR analyses the subset of C-DMRs were removed.

Definition of methylation levels. Throughout this work, we refer to the level of methylation of genomic regions. To compute this level for a given region, we summed the number of sequenced C bases across all cytosines that were called statistically significantly methylated by our pipeline and divided that sum by the number of sequenced bases covering all cytosines in the given region.

Relationship between DNA methylation and mutation. In an attempt to look at the relationship between mutation and DNA methylation, we calculated the weighted average of DNA methylation and mutation rates across all genes. Genes were defined as entries in the TAIR10 reference GFF file having the word "gene" in the feature column. Methylation levels were calculated as described above, and SNP effects were determined using the snpEff tool (http://www.ncbi.nlm.nih.gov/pubmed/22728672) and its *A. thaliana* TAIR10 reference file. We computed two mutation rates, the overall mutation rate and the major effect mutation rate, which we obtained by calculating the fraction of mutations in that gene out of the total number of mutations that were observed in that gene across all accessions. Major effect mutations were defined as mutations that introduced or removed a start or stop codon. The methylation level and mutation rates for each locus were normalized to the maximum value observed at that locus for each measurement type. This normalization yielded measurements on a scale from 0 to 1. We performed a correlation test on these measurements to try and detect a relationship between methylation level and either of the mutation types. As we had no reason to suspect a linear relationship between these variables, we chose to use a Kendall statistic to evaluate the correlation. We detected small but statistically significant relationships between all three of our measurements. Although these results are statistically significant given the small magnitude of the correlation coefficients, we believe that these relationships are at least difficult to interpret but probably not biologically meaningful (Supplementary Table 11).

Enrichment of DMRs in genes and transposons. To determine if CG- and C-DMRs were enriched or depleted in genes or transposons, we performed a binomial test based on the proportions of these features throughout the genome. The results of these tests can be found in Supplementary Table 7.

Linkage disequilibrium/positional association decay analysis. To determine the rate of decay for C-DMRs and CG-DMRs we computed a Pearson correlation coefficient between each pair of DMRs within 10 kb of one another. These coefficients were then separated into 1 kb or 200 bp bins based on the distances between the midpoints of the DMRs. We took the median correlation coefficient of each bin as the rate of decay at a particular distance. In the case of SMPs and SNPs, we used the software package PLINK to determine the association/ linkage disequilibrium between all pairs of sites with a minor allele frequency of 20% and that were within 10 kb of one another. In the case of DMRs, we computed the minor allele frequency by first scoring each accession's DMR as methylated (methylation level $\geq$10%) or unmethylated (methylation level $<$10%). These scores were binned as in the case of DMRs, and the median value of each bin was taken as the decay rate for a particular distance.

DMR saturation analysis. We estimated how close we are to saturating the discovery of DMRs by randomly subsetting our data and calling DMRs on those subsets (Supplementary Fig. 27). For each of the sample sizes, five random subsets were drawn from the samples and run using the same DMR calling pipeline previously outlined. Although the discovery of new CHH-DMRs seems to be saturated, DMRs in the other contexts remain to be found.

mQTL analysis. Given our small sample size, we made several efforts to control for the number of false positives we undoubtedly found. To this end, we only tested DMRs that had at least 75% (114 samples) of their observations present and at least 10% of their observations over a 10% methylation level (that is, what we defined as a methylated allele). Additionally, we only tested phenotypes that had genomic inflation factors (GIFs) between 0.985 and 1.015. To obtain these GIFs, we calculated the 50th percentile of each tested C-DMR's distribution of *P*-values

©2013 Macmillan Publishers Limited. All rights reserved

as well as the 50th percentile of the distribution of *P*-values generated by randomly permuting the phenotypes of 20 randomly chosen C-DMRs 10 times (200 permutations in total). These filtering steps left us with 3,023 C-DMRs and 1,877 CG-DMRs to test. We then randomly sampled 1% of the *P*-values tested and input them to the R package Q-Value[41]. The *P*-value corresponding to a 1% false discovery rate was then used as a cut-off to determine the significance of each association test (we refer to this methodology as the 'Q-value method'). The results for significant SNPs are detailed in Supplementary Table 20. As further validation to ensure that this methodology was working, we compared it to the randomization method outlined elsewhere[42] (we refer to the following methodology as the 'randomization method'). To this end, we randomized the labels in our genotype matrix (that is, so every sample now had genotypes from a different, randomly chosen sample) and ran EMMAX on the DMRs that had passed our quality control thresholds. Specifically, we ran those DMRs that had at least 10% of their DMRs in the 'methylated' state, at least 75% of their observations present, and a GIF between 0.985 and 1.015. For each DMR tested, we attempted to find the largest *P*-value that kept the false discovery rate (FDR) under 1%. In this case, we defined the FDR of a given *P*-value cut-off as the fraction of significant (that is, below the *P*-value cut-off in question) hits found in the randomized set out of the total number of significant hits found in the randomized and non-randomized sets. The results for significant SNPs are detailed in Supplementary Table 21. We found that the methodology employing Q-value discovered fewer mQTL than the randomization method (Supplementary Table 22), but both methods found a similar proportion of cis and trans mQTL (Supplementary Fig. 28). Furthermore, the Q-value results are nearly a perfect subset of the randomization results (~93% overlap). Consequently, to be conservative, we used the SNPs that overlapped in both methodologies for the analysis in the paper. We grouped these significant SNPs into blocks with the following method.

If a significant SNP lies within 10 kb of another significant SNP combine these two SNPs into a block (that is, the block's start and end are now the positions of these two SNPs). Using this block as a starting point, look for other significant SNPs that are within 10 kb of either end of the block. If such SNPs exist, add them to the existing block, update the block ends with the new SNP, and look for significant SNPs within 10 kb of these new block ends. Repeat this procedure until no significant SNPs can be found within 10 kb of the block ends. These blocks are what we refer to as mQTL throughout the paper. To prioritize candidate loci for follow up studies, we have listed all genes (that is, protein-coding genes defined in the file here (ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_gff3/TAIR10_GFF3_genes.gff)) that fall within the mQTL blocks defined by these significant SNPs, the number of significant SNPs that directly overlap these genes, and whether or not they have been implicated in DNA methylation processes (Supplementary Tables 16 and 23). To better address the validity of mQTL that associated with more than one mQTL, we ran the 1,045 C-DMRs with at least one significant mQTL through the MLMM software provided in ref. 32. When evaluating results from this program, we chose the model that minimized the EBIC criterion reported. We used the same *P*-value cut-off given by the Q-value method above to determine which results were significant and collapsed them in the same fashion as mentioned above. We have included the individual results for the significant SNPs in Supplementary Table 24.

Expression of genes containing DMRs. The lists of C-DMRs and CG-DMRs were used to find the overlap between them and a list of protein coding genes (that is, genes with the 'protein-coding gene' descriptor in the TAIR10 reference annotation file found here (ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_gff3/TAIR10_GFF3_genes.gff). We then compared the methylation level of these DMRs with the expression levels of the genes they overlapped. We created boxplots of the expression levels for various methylation levels (for example, the expression values for all genes with a DMR that had a methylation level greater than 0.2 but less than 0.3). All the expression values of a locus were divided by the maximum observed value at that locus, so the expression values plotted are the fraction of the maximum expression level observed at a given locus. It is interesting to note that genes with no C methylation are expressed at a lower level than those that have a methylation level between (but excluding) 0 and 0.1. This dip is due to genes that have no gene body (that is, CG methylation) as has been shown in ref. 20 and is also apparent in these loci (Supplementary Fig. 8). Consequently, we plotted these data again excluding those sites without gene body methylation (that is, 0 now represents loci with no CHG or CHH methylation) and saw the median expression rise to match the median expression level at the 0 to 0.1 level (Supplementary Fig. 9). To make the differences in the medians clearer, we have plotted the median values for the boxplots in Fig. 2k and 2j along with the bootstrap confidence intervals in Supplementary Figs 29 and 30.

Developmental gene expression profiling. Microarray analysis was previously performed for a broad range of developmental stages throughout the plant life cycle[37]. These data were downloaded from http://www.weigelworld.org/resources/microarray/AtGenExpress/AtGE_dev_gcRMA.txt.zip/at_download/file. These lists of loci that are targeted by the RdDM pathway were matched against probe IDs and the resulting information was extracted. Triplicate data for each developmental time point was averaged and then row normalized according to the developmental time point that displayed the highest expression level and then plotted as a heatmap.

Analysis of local sequence variants at C-DMRs overlapping genes. Primer sets were designed and used for PCR amplification of 92 methylated C-DMRs and for amplification of 86 C-DMRs with local mQTL. Individual PCR products were purified with a PCR purification column (Qiagen) and then sequenced with Sanger sequencing technology. All primer sets can be found in Supplementary Tables 13 and 15.

SMP conservation. To get a global look at the diversity of methylation across each chromosome, we binned cytosine positions into 10 kb windows. To examine the conservation of methylation state at cytosines throughout the genome, we computed a score for each site. Any cytosine that had less than five reads covering it was excluded. We used the following formula to estimate the amount of conservation at each site that was missing data from no more than 50 samples: (count(methylated accessions) − count(unmethylated accessions)) / (count(methylated accessions) + count(unmethylated accessions)). This score reaches its maximum value of 1 when all accessions are methylated and a minimum of −1 when all accessions are unmethylated. We computed this score for each site within a bin (Fig. 1a and Supplementary Fig. 5) and then averaged those statistics together. The distributions of these scores are plotted across features in Fig. 1b, d and e.

Genome-wide running correlation of SMP, SNP and C-DMR diversity measures. To evaluate how the correlation between the diversity measures calculated for SMPs, SNPs and C-DMRs changed across the genome, we calculated diversity measures in the same way as in Fig. 3b, but in 100 kb windows offset by 20 kb instead of 500 kb windows offset by 100 kb. We changed the window size and offset in order to generate more points with which to perform correlation tests. First, we calculated the percentiles of all the diversity measures. Next, we performed a Kendall tau correlation test on these percentiles for all windows that started within 500 kb (upstream or downstream) of a genomic coordinate (listed as the Window Center in Supplementary Table 25). The coefficients from these tests as well as their *P*-values are listed in Supplementary Table 25.

38. Li, L. *et al.* Linking photoreceptor excitation to changes in plant architecture. *Genes Dev.* **26,** 785–790 (2012).
39. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
40. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
41. Storey, J. D. A direct approach to false discovery rates. *J. R. Stat. Soc. A* **64,** 479–498 (2002).
42. Breitling, R. *et al.* Genetical genomics: spotlight on QTL hotspots. *PLoS Genet.* **4,** e1000232 (2008).

©2013 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature11973

Membrane potential dynamics of grid cells

Cristina Domnisoru[1,2,3,4], Amina A. Kinkhabwala[1,2,3,4] & David W. Tank[1,2,3,4]

During navigation, grid cells increase their spike rates in firing fields arranged on a markedly regular triangular lattice, whereas their spike timing is often modulated by theta oscillations. Oscillatory interference models of grid cells predict theta amplitude modulations of membrane potential during firing field traversals, whereas competing attractor network models predict slow depolarizing ramps. Here, using *in vivo* whole-cell recordings, we tested these models by directly measuring grid cell intracellular potentials in mice running along linear tracks in virtual reality. Grid cells had large and reproducible ramps of membrane potential depolarization that were the characteristic signature tightly correlated with firing fields. Grid cells also demonstrated intracellular theta oscillations that influenced their spike timing. However, the properties of theta amplitude modulations were not consistent with the view that they determine firing field locations. Our results support cellular and network mechanisms in which grid fields are produced by slow ramps, as in attractor models, whereas theta oscillations control spike timing.

Grid cells[1] in the medial entorhinal cortex (MEC) have been proposed to form the metric needed for mapping space[2]. A widely held view is that sensory cues specifying location allow error correction and set the grid map's orientation in different environments, whereas the periodically repeating grid fields are internally generated by path integration of velocity signals[1,3]. Extensive modelling efforts have produced two broad conceptual frameworks that explain how grid fields arise by velocity integration[4,5]: oscillatory interference models[6–12] and attractor network models[3,13–17].

In oscillatory interference models (Fig. 1a and Supplementary Fig. 1a), grids arise independently in each grid cell. A constant-frequency theta oscillation combines with one or more theta oscillations for which frequency varies with animal velocity to form interference patterns in the membrane voltage[6–12]; the amplitude of intracellular theta is largest 'in-field', with spikes occurring on the peaks of theta cycles. In support, MEC cells exhibit intrinsic oscillations in slices[18,19]; grid cells show theta phase precession[20]; abolishing theta eliminates grid firing[21,22]; and theta- and velocity-modulated cells have been identified[23]. However, grid cells exist in bats without clear theta oscillations[24], and theta oscillations may be too noisy for interference-based integration[25]. The role and importance of theta for grid formation is therefore currently controversial[4,5,26,27] and requires further study.

Alternatively, in attractor network models (Fig. 1b and Supplementary Fig. 1b), grid fields arise from collective dynamics among cells synaptically connected with a specific topology[3,13–17]. Velocity inputs shift the attractor state in the direction of movement. From the perspective of a single grid cell, the moving quasi-stable attractor state of activity generates a slow up and down ramp of synaptically generated depolarization during field traversals. This ramp defines the cell's firing field as spike threshold is crossed (see simulations in Supplementary Fig. 1b). Despite their theoretical appeal, direct evidence in support of these models has remained scarce.

Interference and network models therefore predict that fundamentally different membrane potential signatures—amplitude-modulated theta oscillations and slow up and down ramps—would be the primary drive of firing in grid fields. Here we examine these predictions by direct measurement of the membrane potential of grid cells in mice during navigation in virtual reality[28–30] (Supplementary Fig. 2).

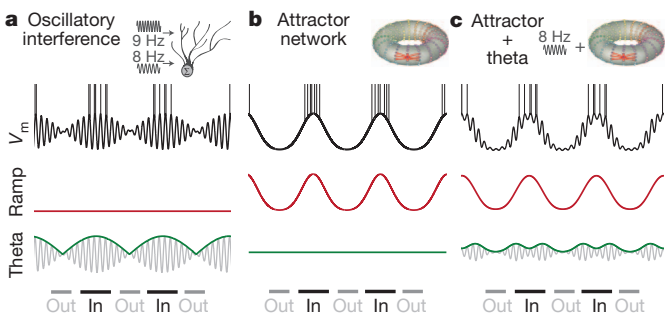


Figure 1 | Schematics of membrane potential predicted by different model families. **a**, Oscillatory interference models. Fields form when theta envelope exceeds threshold; spikes occur on peaks of theta cycles. **b**, Attractor network models. Fields form when ramps exceed threshold, without precise spike timing (attractor schematic reproduced from ref. 3). **c**, Attractor with theta-oscillatory input[17]. Fields form when ramps are high and spikes occur on peaks of theta oscillations that also drive the attractor. In **a**–**c**, the membrane potential (V_m) is decomposed into a low-frequency 'ramp' component (red) and a theta frequency component (grey, envelope in green) during in-field ('In') and out-of-field ('Out') periods.

Grid cells in virtual reality

We first used tetrodes (Supplementary Fig. 3) to record MEC neurons in a real two-dimensional arena[31] (Fig. 2a–c); the same cells were then recorded during navigation along a virtual linear (one-dimensional) track (Fig. 2d–f). Units were identified as grid cells from the recordings in the two-dimensional arena (Supplementary Fig. 4). Grid cells in virtual reality had increased firing rates at multiple locations along the linear track (Fig. 2d–f), consistent with grid cell firing on real linear tracks[20,32]. Firing fields and out-of-field periods were defined using a shuffle test (Supplementary Fig. 5). Grid cell peak firing rates, field width, and field spacing were very similar between virtual and real linear tracks (Supplementary Fig. 6); as in real tracks, grid cell firing rates increased weakly with running speed[33] (Supplementary Fig. 7). These properties indicate that the grid cell circuit operates normally in virtual reality.

[1]Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08544, USA. [2]Bezos Center for Neural Circuit Dynamics, Princeton University, Princeton, New Jersey 08544, USA. [3]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. [4]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA.

©2013 Macmillan Publishers Limited. All rights reserved

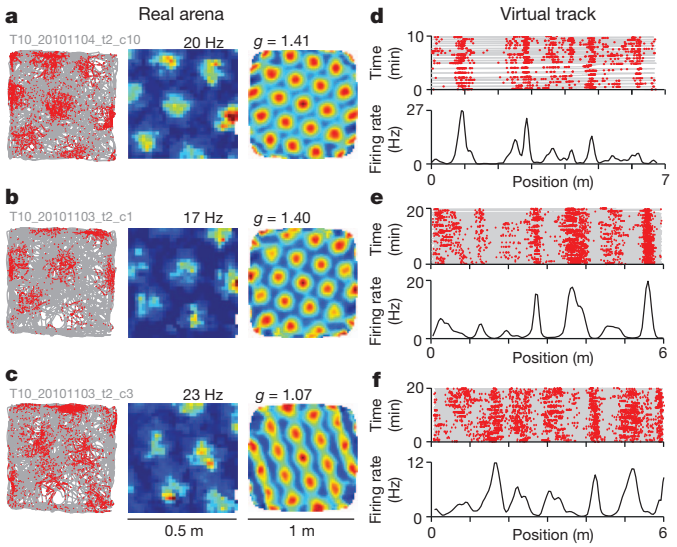


Figure 2 | Tetrode recordings from grid cells in two-dimensional arenas and virtual linear tracks. **a**–**c**, Grid cells identified in a real arena. Trajectory (grey) and spikes (red dots), firing rate (peak rate above), and autocorrelation function are shown for three cells. Grid scores (*g*; Supplementary Methods) are shown above the autocorrelation function. **d**–**f**, Same cells as **a**–**c**, on a virtual linear track. Top: trajectory (grey) and spikes (red dots). Bottom: firing rate; cells **e** and **f** were recorded simultaneously.

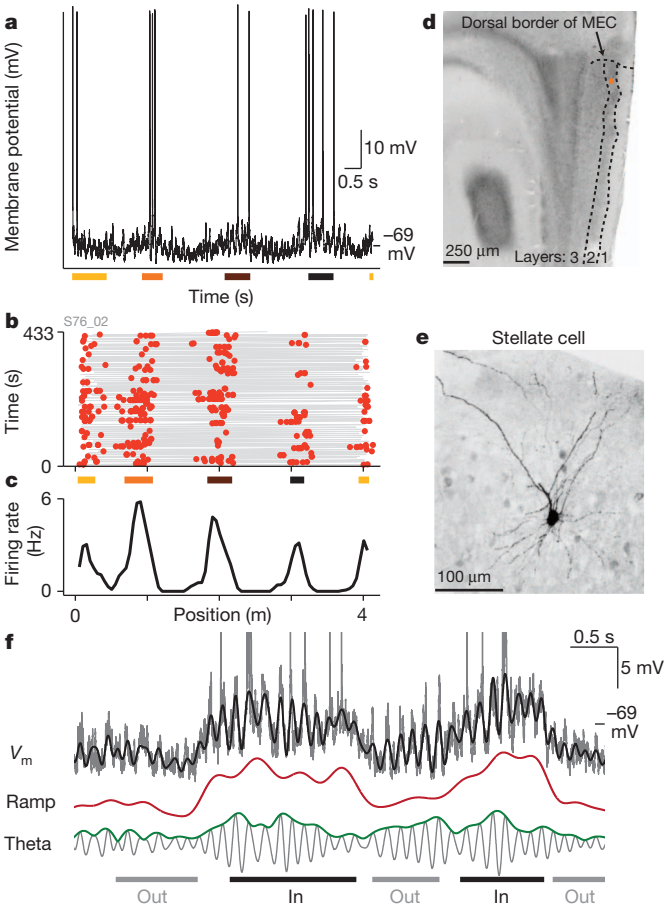


Figure 3 | Whole-cell recordings from grid cells. **a**, Membrane potential versus time during a complete run along the virtual track; coloured bars indicate grid fields. **b**, Trajectory and spikes versus position; coloured bars indicate the same fields as in **a**. **c**, Firing rate. **d**, Soma location (orange dot) in layer 2 of MEC. **e**, Stellate morphology. **f**, The membrane potential (grey, spikes truncated) could be decomposed into ramp (red) and theta components (grey) and was closely approximated by their sum (black overlay).

A classifier was implemented that correctly identified the grid cells within the tetrode data set based only on their linear track firing rates with a high true-positive rate (87%) and a low false-positive rate (13%, Supplementary Figs 8 and 9 and Supplementary Methods). The high success rate allowed us to use this classifier to identify grid cells among our whole-cell recordings from their linear track firing alone.

Whole-cell recordings were performed during navigation as described previously[29], but with improved methods for mechanical stability (mean recording duration 10.3 ± 8.9 min; maximum duration 41.3 min, Supplementary Methods). From 53 recordings, 42 contained enough traversals to measure firing fields. Using our classifier, 27 of these were identified as grid cells (64%, Supplementary Fig. 10). Importantly, the firing field properties of intracellularly recorded grid cells were highly similar to tetrode-recorded grid cells (Fig. 3b, c and Supplementary Fig. 6). A subset of whole-cell recordings had firing fields with highly regular, periodic-like spacing reminiscent of the periodicity of grid cells in two-dimensional environments (Supplementary Fig. 11). The firing rates of highly periodic cells can be viewed as what would be expected if mice ran in a straight line (or 'slice') through adjacent firing fields of a grid cell in a two-dimensional environment, along a grid axis. This was also the case for grid cells that were strikingly aperiodic on a linear track: their irregular firing fields could also be approximated by slices through two-dimensional grid lattices in a direction not parallel to the grid axes (Supplementary Fig. 12).

Because MEC contains grid cells in two hippocampal-projecting layers (2 and 3) that differ markedly in terms of connectivity and cell-type composition[34], biocytin fills were used to recover the morphology and layer of 12 grid cells (Fig. 3d, e and Supplementary Figs 13 and 14). From the 9 grid cells with identified morphology in layer 2, 6 were stellate and 3 were pyramidal, indicating that both cell types could be grid cells in layer 2, whereas all 3 grid cells with known morphology in layer 3 were pyramidal (Fig. 4e, f and Supplementary Fig. 14). Using electrode tracts together with electrophysiological properties, 7 additional grid cells with unknown morphology were localized to layer 2 or 3 (Supplementary Methods).

Grid cell intracellular recordings displayed two prominent membrane potential signatures: theta oscillations and slower depolarizations that appeared to rise and fall during firing fields (Fig. 3a, f). The slow depolarizations and theta oscillations were extracted from the membrane potential after spike removal using complementary digital filters with non-overlapping passbands (Supplementary Fig. 15). Because these were the two primary features, the spike-free membrane potential was always well approximated by their sum (Fig. 3f and Supplementary Fig. 15). This allowed us to study the properties of ramps and theta oscillations separately with respect to action potential firing in grid cells. The ongoing amplitude of theta oscillations was quantified using an estimate of the peak-to-peak theta envelope (Fig. 3f and Supplementary Fig. 15). On the timescale of traversals through sequential firing fields, the slow depolarization component was low between fields, ramping up and down within fields. After subtracting the mean out-of-field value (baseline), we therefore termed the resulting waveform the 'ramp' (Supplementary Fig. 15 and Supplementary Methods).

Grid cell intracellular oscillations

Different grid cells displayed a surprisingly wide range of theta oscillation amplitudes (~2–12 mV theta envelope, Fig. 4a–c). A subset of neurons had striking high-amplitude oscillations, and were termed 'large theta cells' (Fig. 4a, d and Supplementary Figs 16, 17 and 18, part 3). All large theta cells that could be localized were in layer 2 (10 in MEC and 1 in a structure referred to as a 'large dorsal patch'[35]). Theta amplitudes were significantly lower in the remainder of layer 2 'small theta' cells (Fig. 4b, d and Supplementary Fig. 18, part 1) and

©2013 Macmillan Publishers Limited. All rights reserved

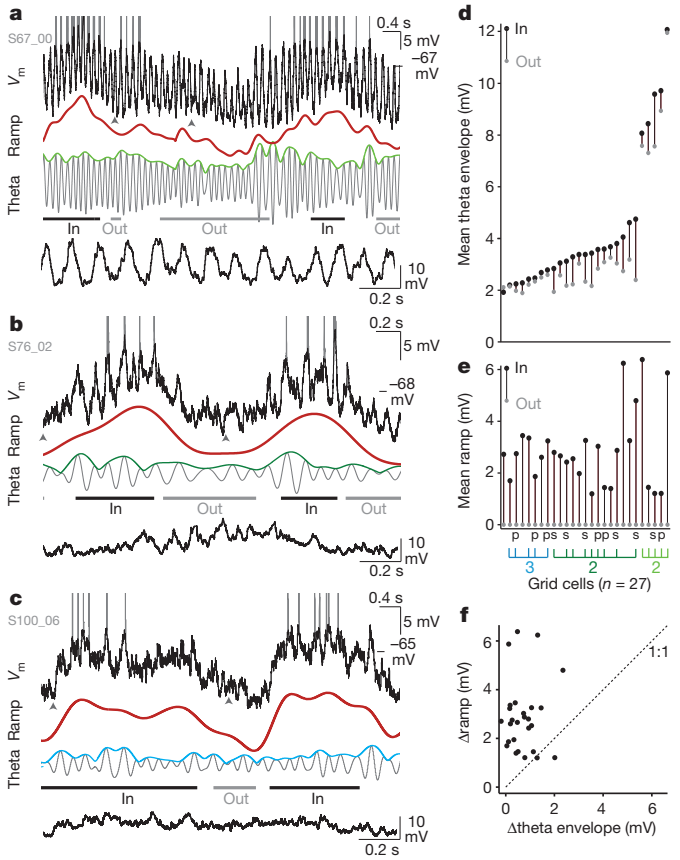


Figure 4 | Grid cells exhibit slow intracellular ramps of depolarization and theta oscillations with variable amplitudes. **a**, Membrane potential of a large theta grid cell (top, spikes truncated) during two field traversals, and its decomposition into the ramp (red) and theta component (grey). Lower trace: expanded view of V_m (between arrowheads) after spike removal. **b**, Same as **a** for a layer 2 small theta cell. **c**, Same as **a** for a layer 3 cell. **d**, Mean theta envelope during in-field (black dots) and out-of-field (grey dots) periods for all grid cells ($n = 27$), sorted by in-field mean. **e**, Same as **d** for the ramp (cells in the same order). Morphology (s, stellate; p, pyramidal) and layer (3 or 2) are indicated underneath; for layer 2, large and small theta cells are light and dark green, respectively. **f**, ΔRamp versus Δtheta envelope for all grid cells (Δramp = mean in-field − mean out-of-field ramp; Δtheta envelope is analogously defined).

in layer 3 recordings[36] (Fig. 4c, d and Supplementary Fig. 18, part 2; one-way analysis of variance (ANOVA), $F_{2,16} = 41.9$, $P < 10^{-6}$; layer 3 or layer 2 small theta versus large theta cells, $P < 10^{-5}$; layer 3 versus layer 2 small theta, not significant, Tukey honestly significant difference (HSD) test). We note that although these groups differ in mean theta amplitude, they did not differ significantly in terms of their ramp amplitudes (interdecile range mean ± s.e.m., in mV: large theta, 7.3 ± 1.6; layer 2 small theta, 6.4 ± 0.6; layer 3, 8.5 ± 1.3, not significant, one-way ANOVA, $F_{2,16} = 1.0$, $P > 0.3$, Tukey HSD test, all $P > 0.3$).

Ramp and theta oscillation changes in firing fields

Attractor network models predict that firing fields occur because of a slowly rising and then falling ramp of synaptic input-induced depolarization; the field is located where the depolarization exceeds spike threshold. Conversely, interference models predict that firing fields occur because the theta envelope increases in the field, causing firing on the depolarizing peaks that exceed spike threshold (Fig. 1a and Supplementary Fig. 1a). We found that the mean ramp increased in all grid cells during in-field compared with out-of-field periods. Theta envelope also increased in all but one cell, but the increases were generally smaller (Fig. 4d, e). To quantify their importance to driving in-field firing on a cell-by-cell basis, we first compared the relative amplitude of Δramp and Δtheta envelope in each neuron and found that Δramp was larger than Δtheta envelope in 25 out of 27 grid cells (Fig. 4f). The ratio between the average Δramp and Δtheta envelope was ~4, indicating that the ramp would be more effective at depolarizing grid cells and driving field formation (mean ± s.e.m., Δramp, 2.9 ± 0.3 mV; Δtheta envelope, 0.72 ± 0.12 mV; Δramp > Δtheta envelope, $t(26) = 7.3$, $P < 10^{-7}$, paired t-test).

The characteristic feature of membrane potential dynamics that forms grid fields should be highly correlated with the firing rate at all times, changing reliably and systematically with position in space. Like the firing rate, this signal should increase as the animal enters a field, reach a peak, then decrease to a minimum between grid fields, and do so consistently on every run through the environment (Fig. 1 and Supplementary Fig. 1). The ramp was strongly correlated with the firing rate (Fig. 5a and Supplementary Fig. 19), both when averaged across runs through the environment (Pearson's correlation coefficient $r = 0.84$) and when the ramp on every run was separately correlated with an estimate of the firing rate on that run ($r = 0.68$, $n = 836$ runs from 27 grid cells). In contrast, the theta envelope was more weakly correlated with the firing rate, sometimes increasing not only in fields but also between them, sometimes decreasing in the middle of grid fields, and typically varying widely from run to run (Fig. 5a and Supplementary Fig. 19; average, $r = 0.46$; run-by-run, $r = 0.22$). Moreover, the ramp was less variable than the theta amplitude when pairs of individual runs were compared to each other. The average pairwise correlation between runs for ramp was significantly larger than for theta envelope ($r = 0.31$ versus 0.08; $t(26) = 7.3$, $P < 10^{-7}$, one-tailed t-test). Together these results demonstrate that the ramp signal tracks the firing rate much more closely than does the theta amplitude.

Because grid cell firing fields occur at specific positions along the track, the characteristic feature of membrane potential that produces firing fields should reliably provide more information about position. Visual inspection of the superimposed runs in Fig. 5a suggests that ramps provide more information, because the theta envelope waveforms are noisier and the averages show less spatial structure. We quantified this by numerically estimating the mutual information between the ramp waveforms and position, and between the theta envelope waveforms and position (Supplementary Methods and Supplementary Fig. 20). In essence, these mutual information estimates quantify how much information (in bits) is gained about location along the track from knowing the ramp or theta envelope voltage. All grid cells recorded demonstrated more information about position along the track in ramps than in theta envelope, with a mean ratio of 2.5 between the two quantities ($n = 27$ grid cells; Fig. 5b).

Because theta was highly variable, we asked whether the increase in theta might be absent in a fraction of individual field traversals, producing counter examples that would be inconsistent with the basic tenets of the theta interference mechanism. In fact, the theta envelope decreased in nearly one-quarter (24%) of field traversals when compared with the out-of-field average, which differs from what interference models predict (examples of in-field firing despite low theta: Fig. 6a and Supplementary Fig. 18). In contrast, the ramp reliably increased in nearly all field traversals (96%). Only field traversals in which the cell spiked are included in this analysis. Furthermore, when the theta envelope increased without a simultaneous increase in the ramp, this was insufficient to drive spiking (no firing despite high theta: Fig. 6b and Supplementary Fig. 18). These counter examples indicated that, even in cells with prominent oscillations, individual field traversals do not necessarily have larger theta envelopes although they nearly always have larger ramps, and that, without the increased ramps, oscillation increases of the magnitudes we observe appear to be insufficient to drive firing.

Beyond individual counter examples, we performed a systematic analysis of the value of the ramp and theta envelope amplitudes when

©2013 Macmillan Publishers Limited. All rights reserved

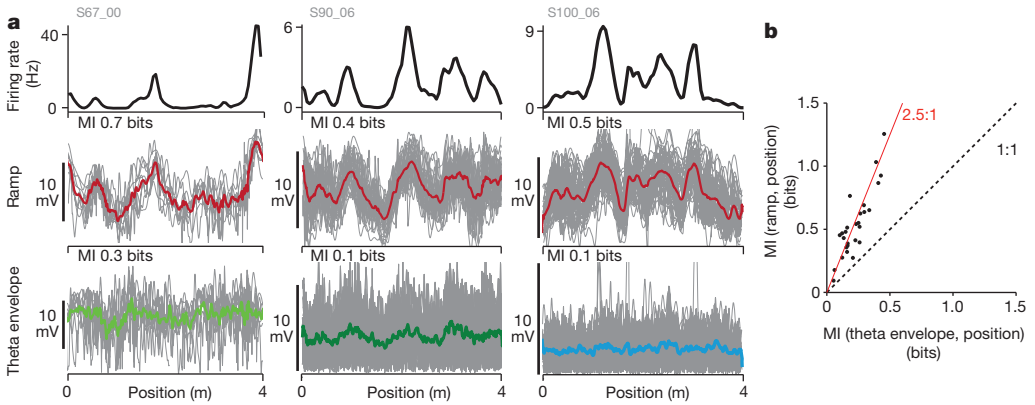


Figure 5 | Ramps contain more information about position. **a**, Three examples of the firing rate (top row), ramp (middle row; individual runs, grey; mean, red) and theta envelope (bottom row; individual runs, grey; mean, green or blue). Left column, large theta ($n = 10$ runs); middle column, layer 2 small theta ($n = 38$ runs); right column, layer 3 ($n = 46$ runs). Ramps show greater run-to-run reproducibility than theta envelopes. **b**, Mutual information (MI; see Supplementary Methods and Supplementary Fig. 20) between ramp or theta envelope and position along the track. Information in ramps is always greater than in theta envelope (points lie above black dotted line that indicates a ratio of 1.0; the mean ratio of MI values is 2.5, indicated by red line).

grid cells spike. For successive 1-ms time intervals during a recording, we determined the associated normalized ramp and theta envelope amplitude and plotted it as a black dot if the cell spiked or as a small grey dot if the cell did not spike at that time point (Fig. 6c). After binning the data in both dimensions, a firing rate estimate was produced at each location and used to plot the effective firing rate for all normalized ramp and theta envelope pairs averaged across cells in each group (Fig. 6d). The analysis directly demonstrates that cells did not spike when the ramp was low (below red lines in Fig. 6c, d; see also Supplementary Fig. 21) regardless of the amplitude of theta. Conversely, when the ramp was high, cells spiked even when theta was low (Fig. 6c, d). On average, spiking was present for the upper ~53% of the range of the ramp and ~93% of the range of theta envelope values of the respective cells. The contributions of ramp and theta envelope to firing fields can be more directly compared by plotting the cumulative spike probability from these distributions as a function of increasing normalized ramp or theta envelope (Fig. 6i). When plotted as a function of ramp amplitude (red lines), there is sharp threshold: above this threshold the cumulative probability for spiking rises quickly to 1, whereas below this value it essentially vanishes. In contrast, the cumulative probability versus theta envelope starts close to 0 and rises immediately, without a threshold, consistent with a cell's ability to spike across the range of available theta envelopes.

Because our recordings were performed in virtual linear environments, in which the firing fields of grid cells are less regularly spaced than in two-dimensional environments, we repeated all analyses for the subset of grid cells that had highly regular, periodic-like firing on linear tracks (7 out of 27 grid cells, Supplementary Methods and Supplementary Fig. 11, part 1). We found no differences in the relative importance of ramps and theta oscillations to field formation in these highly periodic grid cells (Supplementary Fig. 11, part 2). We also repeated all analyses for large theta and small theta cells separately and obtained the same results (analyses not shown). Finally, we defined highly stable firing fields (Supplementary Methods and Supplementary Fig. 10) to examine whether small drifts in firing field location could bias any analysis based on field boundaries, and found no qualitative difference in results.

The fact that grid cells fired when the ramp was high regardless of theta envelope but not when the ramp was low indicated that a ramp increase may be necessary to drive firing in grid cells. Consistent with this, when grid cells failed to fire during individual field traversals ('missed fields'), the ramp was greatly reduced in those field traversals compared to its normal value (mean ± s.e.m. ramp in fields with spikes minus mean ramp in missed fields: 2.64 ± 0.47 mV), confirming that the normal in-field increase in ramp was important for driving in-field firing (Supplementary Fig. 22). Although the theta envelope was also reduced, the reduction was much smaller (0.73 ± 0.24 mV; drop in ramp is larger, $t(23) = 5.0$, $P < 10^{-4}$, one-tailed t-test; Supplementary Fig. 22e, f).

Although our data support the view that ramps are the generator of firing fields, grid cell firing was nevertheless typically theta modulated. Across cell types, spiking occurred preferentially on the peaks of intracellular theta oscillations (Fig. 6e, f; fraction of spikes occurring on the positive half of theta cycles: large theta, 97 ± 2%; layer 2 small theta, 86 ± 3%; layer 3, 74 ± 3%, group means were statistically different, one-way ANOVA, $F_{2,16} = 14.3$, $P < 0.001$, post-hoc Tukey HSD test, all $P < 0.05$). The fact that the average spiking probability was centred on the peaks of intracellular theta oscillations is consistent with a lack of intracellular phase precession. Consistently, we found that, as in place cells[29], grid cells did not exhibit phase precession with respect to intracellular theta oscillations (Fig. 6g, h and Supplementary Fig. 23, part 1). In separate experiments, we performed extracellular recordings from 17 layer 2 cells in MEC where local field potential (LFP) phase precession is expected[20], and found phase precession with respect to the LFP in 7 out of 8 grid cells (from 3 mice, Fig. 6g, h and Supplementary Fig. 23, part 2). To quantify the relative importance of theta and ramp amplitude to precise spike timing, a theta index was calculated for each cell. A regression analysis was performed that estimated this degree of theta modulation as a function of both the amplitude of the cell's ramp and theta envelope. Theta modulation increased with theta envelope amplitude and weakly decreased with ramp amplitude, and the envelope explained a much larger fraction of the variance of theta index than ramp amplitude (Supplementary Fig. 24). This analysis demonstrates that the amplitude of theta oscillations is the primary determinant of precise, theta-paced spike timing.

Discussion

We demonstrated that the membrane potential dynamics of grid cells includes both ramps and theta oscillations. Compared to changes in theta amplitude, ramps are larger, more reliable, more correlated with firing rate, provide more information about position, and unless they increase sufficiently, spiking does not occur. Ramps were present in grid cells in both hippocampal-projecting MEC layers (2 and 3), and in both stellate and pyramidal cells. The presence and importance of ramps is consistent with the primary membrane potential signature of attractor network models of grid cells. Moreover, on average the ramp peak occurred near the middle of the grid field, as predicted by attractor models (Fig. 1 and Supplementary Figs 1 and 25).

©2013 Macmillan Publishers Limited. All rights reserved

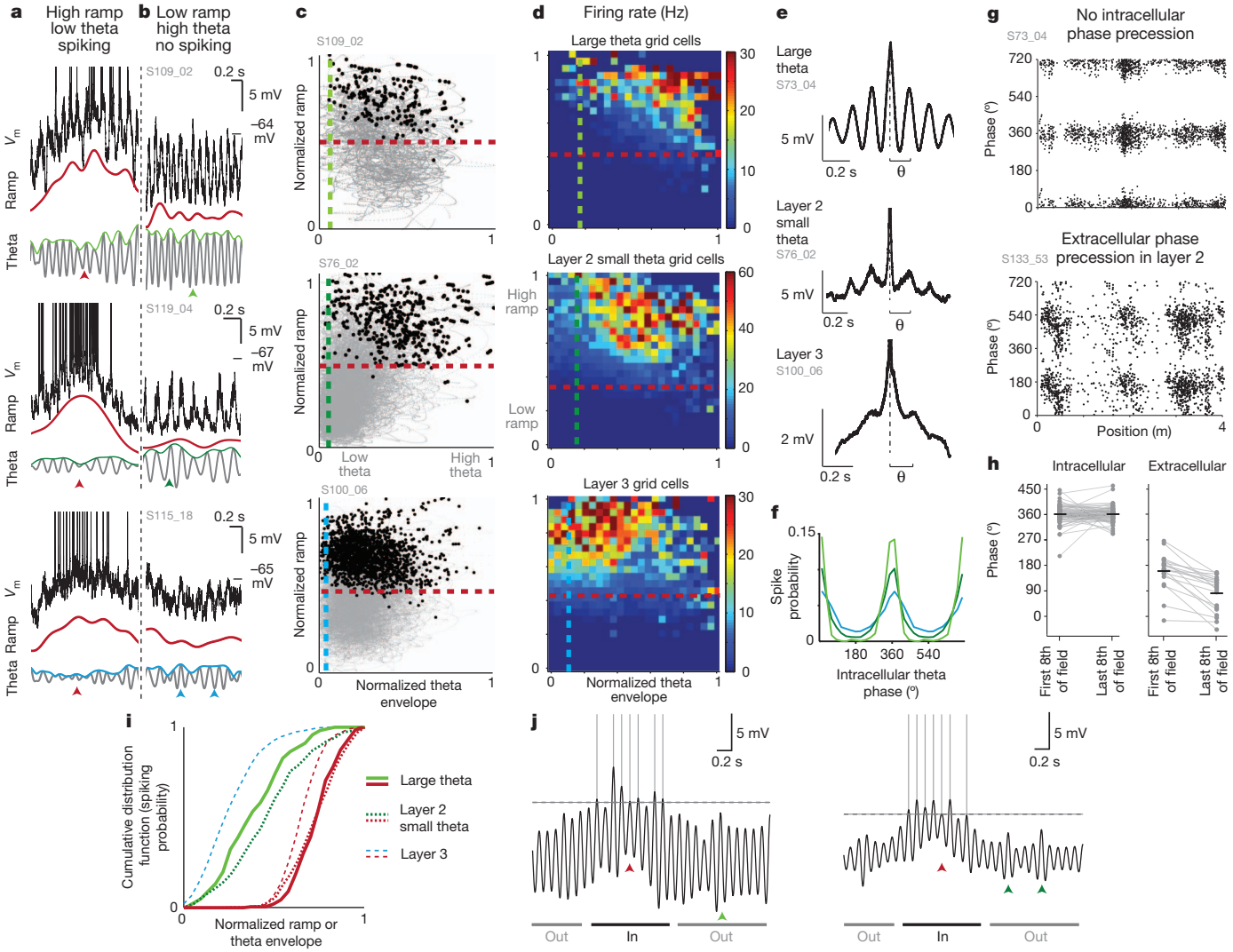


Figure 6 | Ramps, not oscillations, are the primary drive of field formation. **a**, Example field traversals with spikes fired when the ramp (red) is high although theta envelope (green or blue) is low (red arrowheads; top panel, large theta; middle panel, layer 2 small theta; bottom panel, layer 3). **b**, Periods from the same recordings in **a** showing no spiking when theta amplitude is high but the ramp is low (green or blue arrowheads). **c**, Normalized ramp and theta envelope for the whole recording (grey) and during spikes (black) for 3 grid cells (top panel, large theta; middle panel, layer 2 small theta; bottom panel, layer 3). Ninety-nine per cent of spikes occurred at higher ramps than the dashed red lines and higher theta envelope than the dashed green or blue lines. **d**, Average firing rate for the three cell groups in normalized ramp and theta amplitude coordinates. Cells spike for all theta amplitudes, but preferentially above a ramp threshold. **e**, Spike-triggered average (aligned to dashed lines) of the V_m from three cells; spikes occur preferentially on intracellular theta peaks. **f**, Probability of spiking versus intracellular theta phase (large theta, light green; layer 2 small theta, dark green; layer 3, blue). **g**, Top: intracellular theta phase of spikes in a layer 2 large theta grid cell. Bottom: LFP theta phase of spikes in a different layer 2 grid cell. **h**, Mean intracellular (left; $n = 77$ grid fields) and LFP (right; $n = 20$ grid fields) theta phase of spikes fired in the first and last eighth of each field. Horizontal lines are means. **i**, Cumulative distribution function of the spiking probability versus normalized ramp (red) and theta envelope (green or blue) for the cells in **c**. **j**, Schematics of the membrane potential for large theta (left) and layer 2 small theta cells (right, arrowheads same as **a**, **b**). Ramps produce firing fields, whereas spikes are timed by threshold-crossing theta peaks.

We also found increases in ramps and theta amplitude in the spatial firing fields of non-grid cells (Supplementary Fig. 26). This has also been observed in place cells[29]. A suggested explanation for the presence of ramps in place cells in CA1 and non-grid cells in MEC is that they result from summed input from grid cells.

However, in support of interference models, changes of grid scale and theta frequency have been observed dorsoventrally in MEC[32], in novel environments[37], and in HCN1 knockout mice[38]; these properties can arise naturally in interference models by changing the slope of a proposed linear relationship between velocity and theta frequency[9]. Such velocity–frequency relationships have been observed in rats[39]. Here, we also found a weak positive correlation between intracellular theta frequency and running speed (Supplementary Fig. 27). However, as shown by the large standard deviations for each speed (Supplementary Fig. 27), intracellular theta oscillations were very noisy. Without frequent error correction, interference models are thought to tolerate only a small degree of variability in the duration of each theta cycle[10,11,25]. To be consistent with our data, an interference mechanism would have to produce both ramps and noisy oscillations. For example, this could happen if a biophysical process within grid cells could be shown to transform theta-modulated inputs that arrive and interfere in distal dendrites into ramps and noisy oscillations in the soma.

Although the properties of intracellular theta that we measured in grid cells pose challenges to simple forms of interference models, spikes are produced preferentially on the peaks of theta oscillations (Fig. 6e and Supplementary Fig. 21) and a regression analysis demonstrates that theta index, a measure of how modulated spiking is at the theta frequency, increases with increased amplitude of theta oscillations (Supplementary Fig. 24). Thus, intracellular theta

©2013 Macmillan Publishers Limited. All rights reserved

oscillations can control the spike timing of grid cells. When this property of theta oscillations is combined with the sharp threshold of ramp amplitude for firing field spiking (Fig. 6i), a conceptual model emerges for how these two signals combine in a grid cell (Fig. 6j). When the ramp amplitude is low, the probability of spiking is low regardless of theta oscillation amplitude because the membrane potential remains hyperpolarized below spike threshold. As ramp amplitude exceeds spike threshold, a firing field is produced. In the field, the cell is transiently more depolarized during the peaks of the theta cycle than during the troughs, increasing the probability of spike generation on the peaks. This applies to both large theta cells (Fig. 6j, left) and small theta cells (Fig. 6j, right), although larger theta oscillations produce more precise spike timing than smaller oscillations (Fig. 6f).

Our data motivate the consideration of a mechanistic conceptual framework that combines theta oscillations and attractor dynamics[40]. For example, a theta-modulated velocity signal could be used to drive an attractor network. In this view, changing the mapping between animal velocity and theta amplitude and frequency would modulate the drive to the attractor and potentially enlarge or shrink the grids. This could provide an explanation for the observed frequency–velocity relationship in rodents[39]. If velocity inputs are not theta modulated in bats but they are in rodents, this could explain the lack of theta in bat grid cells[24] and its importance in rodents[21,22] while maintaining the same integration mechanism across species. A specific model containing some of these features has been recently proposed in which a theta-modulated velocity signal appropriately shifts the attractor state producing path integration and grid cells with both ramps and theta oscillations[17]. In general, our data are most consistent with an emerging class of models (Fig. 1c and Supplementary Fig. 1c) where the grids arise from attractor dynamics and theta oscillations control the spike timing of grid cells.

METHODS SUMMARY

C57BL/6J mice were trained to run for rewards in real and virtual environments[29] while recordings were obtained from MEC. Tetrode recordings were performed with a microdrive and headplate assembly that could be used interchangeably for navigation in a real two-dimensional arena and a virtual linear track[41]. Whole-cell recordings were obtained as described previously[29], while mice ran along virtual linear tracks. Biocytin fills were used to recover cell morphology and determine soma location. A complete description of the methods is available as Supplementary Information.

Received 19 August 2012; accepted 1 February 2013.
Published online 10 February 2013.

1. Hafting, T., Fyhn, M., Molden, S., Moser, M. B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436,** 801–806 (2005).
2. Moser, E. I., Kropff, E. & Moser, M.-B. Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.* **31,** 69–89 (2008).
3. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. Path integration and the neural basis of the 'cognitive map'. *Nature Rev. Neurosci.* **7,** 663–678 (2006).
4. Giocomo, L. M., Moser, M. B. & Moser, E. I. Computational models of grid cells. *Neuron* **71,** 589–603 (2011).
5. Zilli, E. A. Models of grid cell spatial firing published 2005–2011. *Front. Neural Circ.* **6,** 16 (2012).
6. O'Keefe, J. & Burgess, N. Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus* **15,** 853–866 (2005).
7. Burgess, N., Barry, C. & O'Keefe, J. An oscillatory interference model of grid cell firing. *Hippocampus* **17,** 801–812 (2007).
8. Hasselmo, M. E., Giocomo, L. M. & Zilli, E. A. Grid cell firing may arise from interference of theta frequency membrane potential oscillations in single neurons. *Hippocampus* **17,** 1252–1271 (2007).
9. Burgess, N. Grid cells and theta as oscillatory interference: theory and predictions. *Hippocampus* **18,** 1157–1174 (2008).
10. Giocomo, L. M. & Hasselmo, M. E. Computation by oscillations: Implications of experimental data for theoretical models of grid cells. *Hippocampus* **18,** 1186–1199 (2008).
11. Hasselmo, M. E. Grid cell mechanisms and function: contributions of entorhinal persistent spiking and phase resetting. *Hippocampus* **18,** 1213–1229 (2008).
12. Blair, H. T., Welday, A. C. & Zhang, K. Scale-invariant memory representations emerge from moiré interference between grid fields that produce theta oscillations: a computational model. *J. Neurosci.* **27,** 3211–3229 (2007).
13. Samsonovich, A. & McNaughton, B. L. Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.* **17,** 5900–5920 (1997).
14. Fuhs, M. C. & Touretzky, D. S. A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.* **26,** 4266–4276 (2006).
15. Guanella, A., Kiper, D. & Verschure, P. A model of grid cells based on a twisted torus topology. *Int. J. Neural Syst.* **17,** 231–240 (2007).
16. Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLOS Comput. Biol.* **5,** e1000291 (2009).
17. Navratilova, Z., Giocomo, L. M., Fellous, J. M., Hasselmo, M. E. & McNaughton, B. L. Phase precession and variable spatial scaling in a periodic attractor map model of medial entorhinal grid cells with realistic after-spike dynamics. *Hippocampus* **22,** 772–789 (2011).
18. Alonso, A. & Llinas, R. R. Subthreshold Na^+-dependent theta-like rhythmicity in stellate cells of entorhinal cortex layer II. *Nature* **342,** 175–177 (1989).
19. Giocomo, L. M., Zilli, E. A., Fransen, E. & Hasselmo, M. E. Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science* **315,** 1719–1722 (2007).
20. Hafting, T., Fyhn, M., Bonnevie, T., Moser, M. B. & Moser, E. I. Hippocampus-independent phase precession in entorhinal grid cells. *Nature* **453,** 1248–1252 (2008).
21. Brandon, M. P. *et al.* Reduction of theta rhythm dissociates grid cell spatial periodicity from directional tuning. *Science* **332,** 595–599 (2011).
22. Koenig, J., Linder, A. N., Leutgeb, J. K. & Leutgeb, S. The spatial periodicity of grid cells is not sustained during reduced theta oscillations. *Science* **332,** 592–595 (2011).
23. Welday, A. C., Shlifer, I. G., Bloom, M. L., Zhang, K. & Blair, H. T. Cosine directional tuning of theta cell burst frequencies: evidence for spatial coding by oscillatory interference. *J. Neurosci.* **31,** 16157–16176 (2011).
24. Yartsev, M. M., Witter, M. P. & Ulanovsky, N. Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature* **479,** 103–107 (2011).
25. Welinder, P. E., Burak, Y. & Fiete, I. R. Grid cells: the position code, neural network models of activity, and the problem of learning. *Hippocampus* **18,** 1283–1300 (2008).
26. Giocomo, L. M. & Moser, E. I. Spatial representation: maps in a temporal void. *Curr. Biol.* **21,** R962–R964 (2011).
27. Barry, C., Bush, D., O'Keefe, J. & Burgess, N. Models of grid cells and theta oscillations. *Nature* **488,** http://dx.doi.org/nature11276 (2012); Reply **488,** http://dx.doi.org/nature11277 (2012).
28. Holscher, C., Schnee, A., Dahmen, H., Setia, L. & Mallot, H. A. Rats are able to navigate in virtual environments. *J. Exp. Biol.* **208,** 561–569 (2005).
29. Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461,** 941–946 (2009).
30. Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L. & Tank, D. W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neurosci.* **13,** 1433–1440 (2010).
31. Fyhn, M., Hafting, T., Witter, M. P., Moser, E. I. & Moser, M. B. Grid cells in mice. *Hippocampus* **18,** 1230–1238 (2008).
32. Brun, V. H. *et al.* Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus* **18,** 1200–1212 (2008).
33. Sargolini, F. *et al.* Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science* **312,** 758–762 (2006).
34. Witter, M. P. & Moser, E. I. Spatial representation and the architecture of the entorhinal cortex. *Trends Neurosci.* **29,** 671–678 (2006).
35. Burgalossi, A. *et al.* Microcircuits of functionally identified neurons in the rat medial entorhinal cortex. *Neuron* **70,** 773–786 (2011).
36. Quilichini, P., Sirota, A. & Buzsaki, G. Intrinsic circuit organization and theta-gamma oscillation dynamics in the entorhinal cortex of the rat. *J. Neurosci.* **30,** 11128–11142 (2010).
37. Barry, C., Heys, J. G. & Hasselmo, M. E. Possible role of acetylcholine in regulating spatial novelty effects on theta rhythm and grid cells. *Front. Neural Circ.* **6,** 5 (2012).
38. Giocomo, L. M. *et al.* Grid cells use HCN1 channels for spatial scaling. *Cell* **147,** 1159–1170 (2011).
39. Jeewajee, A., Barry, C., O'Keefe, J. & Burgess, N. Grid cells and theta as oscillatory interference: electrophysiological data from freely moving rats. *Hippocampus* **18,** 1175–1185 (2008).
40. Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J. & McNaughton, B. L. Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus* **6,** 271–280 (1996).
41. Kinkhabwala, A. A. & Tank, D. W. *Spatial Patterning of Grid Cell Firing in Virtual Reality Environments.* Program No. 100.14/KKK30 (Society for Neuroscience, 2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Harvey, R. Low, A. Miri, S. Lewallen, J. Rickgauer, D. Little, D. Barson, D. Aronov, W. Bialek, I. Fiete and G. Buzsaki for helpful discussions. This work was supported by NINDS grant 5RC1NS068148-02 and 1R37NS081242-01, NIMH grant 5R01MH083686-04, NIH Postdoctoral Fellowship grant F32NS070514-01A1 (A.A.K.), and an NSF Graduate Research Fellowship (C.D.).

Author Contributions C.D. performed whole-cell recording experiments and histological identification. A.A.K. performed tetrode experiments; A.A.K. and D.W.T. designed the system for measuring grid cell activity in virtual reality; C.D. analysed the data with strategy and methods contributions from A.A.K. and D.W.T.; C.D. and D.W.T. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.W.T. (dwtank@princeton.edu).

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature11902

Observation of quantum state collapse and revival due to the single-photon Kerr effect

Gerhard Kirchmair[1], Brian Vlastakis[1], Zaki Leghtas[2], Simon E. Nigg[1], Hanhee Paik[1], Eran Ginossar[3], Mazyar Mirrahimi[1,2], Luigi Frunzio[1], S. M. Girvin[1] & R. J. Schoelkopf[1]

To create and manipulate non-classical states of light for quantum information protocols, a strong, nonlinear interaction at the single-photon level is required. One approach to the generation of suitable interactions is to couple photons to atoms, as in the strong coupling regime of cavity quantum electrodynamic systems[1,2]. In these systems, however, the quantum state of the light is only indirectly controlled by manipulating the atoms[3]. A direct photon–photon interaction occurs in so-called Kerr media, which typically induce only weak nonlinearity at the cost of significant loss. So far, it has not been possible to reach the single-photon Kerr regime, in which the interaction strength between individual photons exceeds the loss rate. Here, using a three-dimensional circuit quantum electrodynamic architecture[4], we engineer an artificial Kerr medium that enters this regime and allows the observation of new quantum effects. We realize a gedanken experiment[5] in which the collapse and revival of a coherent state can be observed. This time evolution is a consequence of the quantization of the light field in the cavity and the nonlinear interaction between individual photons. During the evolution, non-classical superpositions of coherent states (that is, multi-component 'Schrödinger cat' states) are formed. We visualize this evolution by measuring the Husimi *Q* function and confirm the non-classical properties of these transient states by cavity state tomography. The ability to create and manipulate superpositions of coherent states in such a high-quality-factor photon mode opens perspectives for combining the physics of continuous variables[6] with superconducting circuits. The single-photon Kerr effect could be used in quantum non-demolition measurement of photons[7], single-photon generation[8], autonomous quantum feedback schemes[9] and quantum logic operations[10].

A material whose refractive index depends on the intensity of the light field is called a Kerr medium. A light beam travelling through such a material acquires a phase shift $\phi_{\mathrm{Kerr}} = K\tau I$ (ref. 11), where K is the Kerr constant, τ is the interaction time of the light field with the material, and I is the intensity of the beam. The Kerr effect is a widely used phenomenon in nonlinear quantum optics, and has been successfully used to generate quadrature and amplitude squeezed states[12], parametrically convert frequencies[13], and create ultra-fast pulses[14]. In the field of quantum optics with microwave circuits, the direct analogue of the Kerr effect is naturally created by the nonlinear inductance of a Josephson junction (specifically the $\phi^4 \propto (b + b^\dagger)^4$ term in the Taylor expansion of $\cos\phi$ in the Josephson energy relation, where ϕ is the superconducting phase difference across the junction, and b and $b^\dagger$ are the corresponding ladder operators)[15,16]. This effect has been used to create Josephson parametric amplifiers[17] and to generate squeezing of microwave fields[18]. However, in both the microwave and optical domains, most experiments use the Kerr nonlinearity in a semi-classical regime, where the quantization of the light field does not play a crucial role. The Kerr effect for a quantized mode of light with frequency ω_c can be described by the normal ordered Hamiltonian, $H_{\mathrm{Kerr}} = \hbar\omega_c a^\dagger a - \hbar\frac{K}{2}a^\dagger a^\dagger aa$, with K the Kerr frequency shift per photon[2,16] and $a/a\dagger$ the ladder operators. The average phase shift per photon is again given by $\phi_{\mathrm{Kerr}} = K/\kappa$, with κ the decay rate of the photon mode. Typical Kerr effects are so small that they are not visible on the single-photon level because $\kappa > K$. Applications which require K much bigger than κ include the realization of quantum logic operations[10], schemes for continuous variable quantum information protocols[6] and quantum non-demolition measurements of propagating photons[7].

The Kerr nonlinearity of a Josephson junction is also routinely used to create superconducting qubits. In this case, a circuit is engineered with such a strong anharmonicity that it can be considered as a two-level system. By combining a qubit with linear resonators, one realizes the analogue of strong coupling cavity quantum electrodynamics (QED), known as circuit QED[1]. This can be used to protect the qubit from spontaneous emission, manipulate and read out its quantum state and couple it to other qubits. One consequence of coupling any resonator to a qubit is that the resonator always acquires a finite anharmonicity K, becoming a Kerr medium itself[15]. Here we have designed K large enough ($K/\kappa > 30$) to be well within the single-photon Kerr regime. At the same time, the nonlinearity is small enough that short pulses ($t_{\mathrm{pulse}} \approx 10\ \mathrm{ns} \ll 1/K$) applied to the resonator create a coherent state, allowing us to conveniently access the large Hilbert space of the oscillator. Only a few experiments have previously come close to the limit where $K/\kappa \approx 1$ while still maintaining the ability to create coherent states[19–21]. So far no experiment has been able to satisfy both conditions at the same time so as to observe the collapse and revival of a coherent state due to the Kerr effect.

As a first demonstration within this single-photon Kerr regime, we realize a proposal[5] for creating photonic Schrödinger cat states. Specifically, we generate coherent states with a mean photon number of up to four photons and measure the Husimi Q function of the resonator state using a new experimental measurement protocol. We then show the high quality of the Kerr resonator by measuring the time evolution of the collapse and revival of a coherent state. During the evolution, highly non-classical superpositions of coherent states, that is, multi-component Schrödinger cat states, are formed, which show the coherent nature of the effect. This revival, in contrast to the Jaynes–Cummings revival of Rabi oscillations of a qubit induced by a coherent state[2,22], is the revival of a coherent state in a resonator. An analogous effect was indirectly observed in early experiments with a condensate of bosonic atoms in an optical lattice[23]. Additionally, we confirm the non-classical properties of the transient states by performing cavity state tomography.

We experimentally realize a highly coherent Kerr medium by coupling a superconducting 'vertical' transmon qubit to two three-dimensional waveguide cavities, as shown in Fig. 1a. This design is based on a recently developed three-dimensional circuit QED architecture[4]. The two halves of the cavities are machined out of a block of superconducting aluminium (alloy 6061 T6). Both cavities have a total quality

[1]Departments of Physics and Applied Physics, Yale University, New Haven, Connecticut 06511, USA. [2]INRIA Paris-Rocquencourt, Domaine de Voluceau, BP 105, 78153 Le Chesnay Cedex, France. [3]Department of Physics and Advanced Technology Institute, University of Surrey, Guildford, Surrey GU2 7XH, UK.

©2013 Macmillan Publishers Limited. All rights reserved

factor of about one million, limited by internal losses, corresponding to a single-photon decay rate $\kappa/2\pi = 10$ kHz. The vertical transmon consists of a single Josephson junction embedded in a transmission line structure, which couples the junction to both cavities. The observed transition frequency of the qubit is $\omega_q/2\pi = 7.8503$ GHz and its anharmonicity is $K_q/2\pi = (\omega_{ge} - \omega_{ef})/2\pi = 73.4$ MHz using the standard convention for labelling from lowest to highest energy level in the qubit as (g,e,f,h,…) (see Supplementary Information). The energy relaxation time of the qubit is $T_1 = 10$ μs with a Ramsey time $T_2^* = 8$ μs. The qubit is used to interrogate the state of the storage cavity, which acts as a Kerr medium. The other cavity is used to read out the state of the qubit after the interrogation, similarly to ref. 24.

The analysis of the distributed stripline elements and the cavity electrodynamics can be performed using finite-element calculations for the actual geometry. Combined with 'black-box' circuit quantization[15], one can derive dressed frequencies, couplings and anharmonicities with good relative accuracy (see Supplementary Information). For the purposes of the experiments discussed here, the coupling of the qubit to the storage resonator, in the strong dispersive limit of circuit QED, is well described by the Hamiltonian

$$\frac{H}{\hbar} = \frac{\omega_q}{2}\sigma_z - \frac{\chi}{2}a^\dagger a\sigma_z + \left(\omega_c - \frac{\chi}{2}\right)a^\dagger a - \frac{K}{2}a^\dagger a^\dagger aa \qquad (1)$$

taking into account only the lowest two energy levels of the qubit. The operators $a^\dagger/a$ are the usual raising/lowering operators for the harmonic oscillator and σ_z is the Pauli operator. In this description, we completely omit the measurement cavity because it is only used for reading out the state of the qubit and otherwise stays in its ground state. The energy level diagram described by the Hamiltonian given in equation (1) can be seen in Fig. 1b. The second term on the right-hand side of equation (1) is the state-dependent shift per photon $\chi/2\pi = 9.4$ MHz of the qubit transition frequency. The last two terms on the right-hand side of equation (1) describe the cavity as an anharmonic oscillator with a dressed resonance frequency ω_c and a nonlinearity $K/2\pi = 325$ kHz which is given by $K \approx \chi^2/4K_q$ (ref. 15). All interaction strengths in the above Hamiltonian are at least one order of magnitude bigger than any decoherence rate in the system.

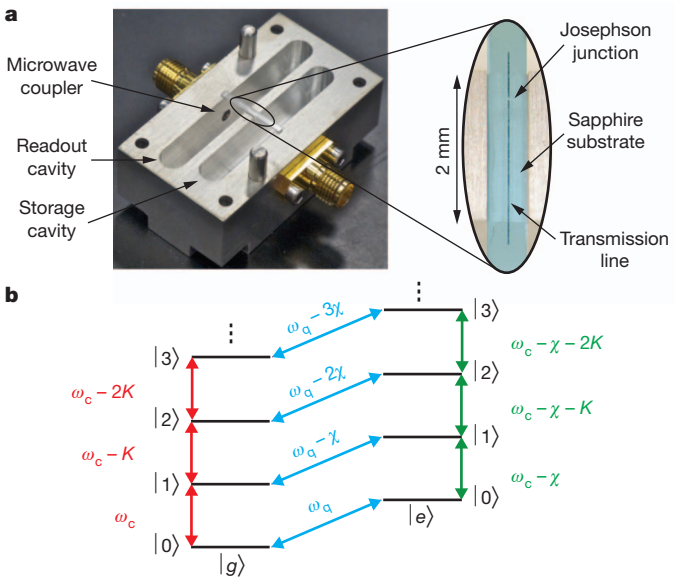


Figure 1 | Device layout and energy level diagram of the two-cavity, one-qubit device. a, Photograph of one half of two aluminium (alloy 6061) waveguide cavities coupled to a vertical transmon qubit. Right, a magnified view of the indicated area of the photograph, showing a detail of the qubit, fabricated on a c-plane sapphire substrate 1.4 mm wide, 15 mm long and 430 μm thick. The coupling strength of the qubit is determined by the length of the stripline coupling antenna which extends into each cavity. The upper cavity, with a resonant frequency of $\omega_m/2\pi = 8.2564$ GHz, is used for qubit readout, and the lower cavity, with a frequency of $\omega_c/2\pi = 9.2747$ GHz, is used to store and manipulate quantum states. **b**, Combined energy level diagram of the qubit coupled to the storage cavity. The qubit states are denoted as $|g\rangle$ and $|e\rangle$, respectively, while the cavity states are labelled as $|n\rangle$, with n the number of photons in the cavity. Each photon in the cavity reduces the qubit transition frequency by χ. Equivalently, exciting the qubit reduces the cavity transition frequency by χ. The energy levels of the cavity are not evenly spaced owing to the induced Kerr anharmonicity, $K/2\pi = 325$ kHz.

To visualize and understand the evolution of the resonator state, we measure the Husimi Q function Q_0 in a space spanned by the expectation value of the dimensionless field quadratures Re(α) and Im(α). Q_0 is defined as the modulus squared of the overlap of the resonator state $|\Psi\rangle$ with a coherent state $|\alpha\rangle$ by $Q_0(\alpha) = \frac{1}{\pi}|\langle\alpha|\Psi\rangle|^2$. Alternatively, we can write Q_0 using the displacement operator $D_\alpha = e^{\alpha a^\dagger - \alpha^* a}$ (note that $D_\alpha^\dagger = D_{-\alpha}$) as $Q_0(\alpha) = \frac{1}{\pi}|\langle 0|D_{-\alpha}|\Psi\rangle|^2$, which describes the actual measurement procedure used in the experiment. The sequence to measure Q_0 can be seen in Fig. 2a. The initial displacement, D_β, creates a coherent state $|\Psi\rangle = |\beta\rangle$ in the cavity, whose Q_0 is given by a Gaussian, $\frac{1}{\pi}e^{-|\alpha-\beta|^2}$. After a variable waiting time t, we measure $Q_0(\alpha)$ by displacing the cavity state by $-\alpha$ and determine the overlap of the resulting wavefunction with the cavity ground state. The population of the cavity ground state can be measured by applying a photon number state selective π pulse, $X_\pi^{n=0}$, to the qubit (see Supplementary Information), similarly to ref. 25. The qubit is excited if and only if the cavity is in the $n = 0$ Fock state, n being the photon number, after the analysis displacement. This scheme allows us to determine $Q_0(\alpha)$ of the resonator up to experimental imperfections (see Supplementary Information for details). Applying π pulses to the qubit conditioned on other photon numbers, X_π^n, allows us to measure the overlap of the displaced state with any Fock state n, which we will call the generalized Q functions $Q_n(\alpha) = \frac{1}{\pi}|\langle n|D_{-\alpha}|\Psi\rangle|^2$. In essence, we can ask the question 'are there n photons in the resonator?', using photon number state selective pulses[24]. To test the analysis protocol, we measured Q_0 and Q_1 of the cavity in the ground state, Fig. 2b–e, by omitting the first displacement pulse of the sequence given in Fig. 2a.

Using this method, we can follow the time evolution of a coherent state in the presence of the Kerr effect. In the experiment, we prepare a coherent state with an average photon number $|\beta|^2 = \bar{n} = 4$ using a microwave pulse[21] to displace the cavity state. We then measure Q_0 for different delays between the preparation and analysis pulses. A comparison of the theoretical evolution of the coherent state and the measured evolution can be seen in Fig. 3. The time evolution of the state is described by considering the action of the Kerr Hamiltonian H_{Kerr} on a coherent state $|\beta\rangle$ in the cavity[5,26]. In the rotating frame of the harmonic oscillator, with the qubit in the ground state, we can write:

$$|\Psi(t)\rangle = e^{i\frac{K}{2}(a^\dagger a)^2 t}|\beta\rangle = e^{-|\beta|^2/2}\sum_n \frac{\beta^n}{\sqrt{n!}}e^{i\frac{K}{2}n^2}|n\rangle \qquad (2)$$

For short times, the nonlinear phase evolution of the Fock states $|n\rangle$ is closely approximated by a rotation of the state with an angle $\phi_{\mathrm{Kerr}} = Kt(|\beta|^2 + 1/2)$ with respect to the frame rotating at ω_c. The onset of this rotation can be seen in Fig. 3a, which is taken at the minimal waiting time of 15 ns between the two displacement pulses. Because of this waiting time, the state rotates under the influence of the Kerr effect from $\beta = 2$ to $\beta e^{i\phi_{\mathrm{Kerr}}} = 2e^{i0.13}$. For longer times we can see how the state rotates further and spreads out on a circle (Fig. 3b, c). This spreading can be simply understood in a semi-classical picture, in which the amplitude components in the coherent state further away from the origin evolve with a higher angular velocity given by the n^2 dependence of the Kerr effect. Complete phase collapse is reached at a time when the phase dispersion across the width of the photon number distribution corresponds to ~π, which can be estimated as $T_{\mathrm{col}} = \frac{\pi}{2\sqrt{\bar{n}}K}$ (ref. 2). For our system, the complete phase collapse

©2013 Macmillan Publishers Limited. All rights reserved

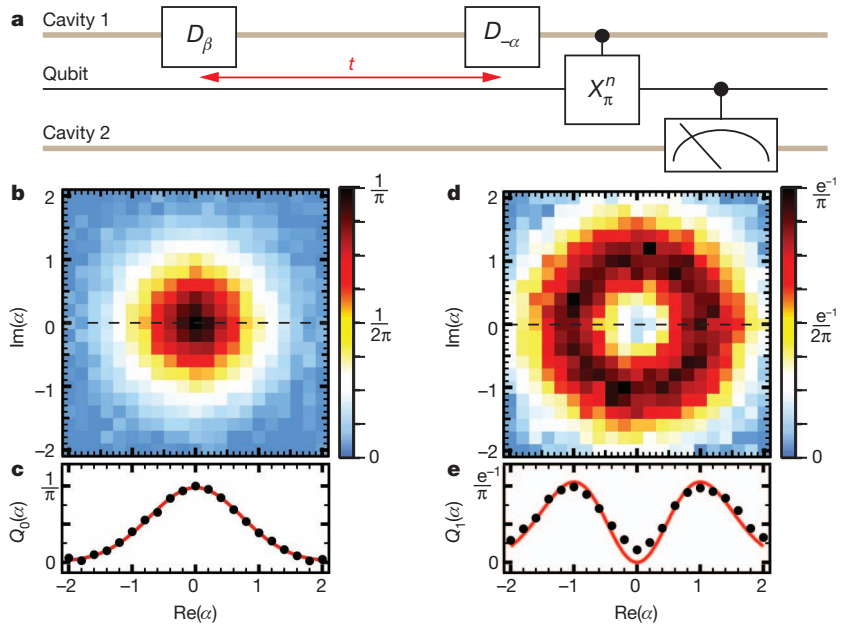


Figure 2 | Technique for measuring the generalized Husimi *Q* functions. **a**, The experimental pulse sequence consists of a 10-ns displacement pulse D_β, which creates a coherent state in the cavity. After a variable waiting time, t, we analyse the state in the cavity by displacing the cavity by $-\alpha$ followed by a π pulse on the qubit conditioned on having n photons in the cavity. In this way we can measure the generalized Q functions, $Q_n(\alpha)=\frac{1}{\pi}|\langle n|D_{-\alpha}|\Psi\rangle|^2$, which are projections of the displaced wavefunction onto the Fock states $|n\rangle$, where $Q_0(\alpha)$ is the Husimi Q function. **b**, Density plot of Q_0 of the ground state with the respective colour scale to the right. We measure the Q function at 441 different analysis displacements α. The π pulse on the qubit $X_\pi^{n=0}$ is conditioned on having no photons in the cavity after the analysis displacement. **c**, Linecut of Q_0 of the ground state along $\mathrm{Im}(\alpha)=0$. The red line is a plot of the theory, a Gaussian given by $\frac{1}{\pi}e^{-|\alpha|^2}$. A fit to the data with a Gauss function $\frac{1}{\pi}e^{-I^2/2\sigma^2}$ results in $2\sigma^2=1.03\pm0.02$, which is consistent with the expected width. **d**, Density plot of Q_1 of the ground state with the respective colour scale to the right. In this case the π pulse on the qubit $X_\pi^{n=1}$ is conditioned on having one photon in the cavity after the analysis displacement. **e**, Linecut of Q_1 of the ground state along $\mathrm{Im}(\alpha)=0$. The red line is a plot of the theory given by a Poisson distribution $\frac{1}{\pi}|\alpha|^2e^{-|\alpha|^2}$.

happens at 385 ns. Figure 3b shows that a Kerr medium can be used as a resource to generate squeezed states[26]. The state is squeezed along the Re(α) quadrature with a width of 0.88(1), as predicted from theory. The maximum squeezing occurs at $t=58$ ns with a width of Q_0 of 0.87.

After the complete phase collapse, structure re-emerges in the form of multi-component superpositions of coherent states (Fig. 3d–f) at times which are integer submultiples of the complete revival $T_{\mathrm{rev}}=\frac{2\pi}{K}$, Fig. 3h. The revivals, periodically appearing every T_{rev}, can be understood by noting that $e^{\frac{iK}{2}n^2t}=(-1)^{n^2}$ for $t=T_{\mathrm{rev}}$. The cavity state is then given by $|\Psi(T_{\mathrm{rev}})\rangle=|-\beta\rangle$. At this time, we get a complete state revival to a coherent state with opposite phase. For $t=T_{\mathrm{rev}}/q$, with q an integer larger than 1, we can write the state of the oscillator as a superposition of q coherent states[2]:

$$\left|\Psi\left(\frac{T_{\mathrm{rev}}}{q}\right)\right\rangle=\frac{1}{2q}\sum_{p=0}^{2q-1}\sum_{k=0}^{2q-1}e^{ik(k-p)\frac{\pi}{q}}\left|\beta\,e^{ip\frac{\pi}{q}}\right\rangle \qquad (3)$$

For $q=2$, we get the two-component Schrödinger cat state, similar to the cat states created in refs 22 and 27. To distinguish the q components of a cat state, the coherent states have to be separated by more than twice their width on a circle with a radius given by the initial displacement. In other words, the coherent states have to be quasi-orthogonal. This means that for a displacement of $|\beta|=2$, the maximum number of coherent states that can be distinguished is 4.

In Fig. 3g, we can see how the state again completely dephases shortly before the coherent revival in Fig. 3h after $t=3{,}065$ ns. After this time, we get a state with amplitude $|\beta|=1.78(2)$, which fits to the expected decay of the resonator state. The theoretical plots for Fig. 3 were simulated by solving a master equation using the decay rate $\kappa/2\pi=10$ kHz of the resonator and introducing a small detuning of 5 kHz of our drive from the resonator frequency ω_c. The time evolution of the state in the experiment agrees well with the theory. The hazy ring that can be seen in theory and experiment in Fig. 3h is produced by cavity decay. The evolution of the state from 0 to 6.05 μs over 50 frames, including two revivals, can be seen in the Supplementary Video.

To get a more quantitative comparison of experiment and theory, we need to determine the quantum state of the resonator. Although in principle one can reconstruct the cavity wavefunction from the measured Husimi Q function, in practice there is important information, such as the interference fringes between coherent state superpositions, which is exponentially suppressed as the separation of the coherent states increases[2]. This makes it hard to distinguish a mixture of coherent states from a coherent superposition in an experiment due to a finite signal to noise ratio. An experimentally more practical way to determine the quantum state of a resonator is, for example, to reconstruct its Wigner function, as this emphasizes the interference fringes. The Wigner function of a cavity[2] has been determined by measuring the parity of the resonator state. More recently, the state of an itinerant microwave field has been determined by measuring the statistical moments of the field operator[28]. Here we use a modified technique, based on earlier work with ion traps and microwave circuits[3,29], to determine the density matrix of the resonator. The main difference is our ability to directly measure $Q_n(\alpha)$, which allows for a simple and efficient measurement and reconstruction of the density matrix of the resonator by using a least square fit to each Q_n (see Supplementary Information). Using this density matrix, we then calculate and plot the Wigner function to show the interference fringes, highlighting the quantum features of the resonator state.

In Fig. 4 we compare the experimentally obtained Wigner functions to a simulation at three different times during the state evolution. The times ($t=2\pi/2K$, $2\pi/3K$ and $2\pi/4K$) were selected such that the Wigner functions correspond to two-, three- and four-component cat states. The simulation was again done by solving a master equation that includes the decay of the cavity. The fidelity $F=\langle\Psi_{\mathrm{id}}|\rho_{\mathrm{m}}|\Psi_{\mathrm{id}}\rangle$ of

©2013 Macmillan Publishers Limited. All rights reserved

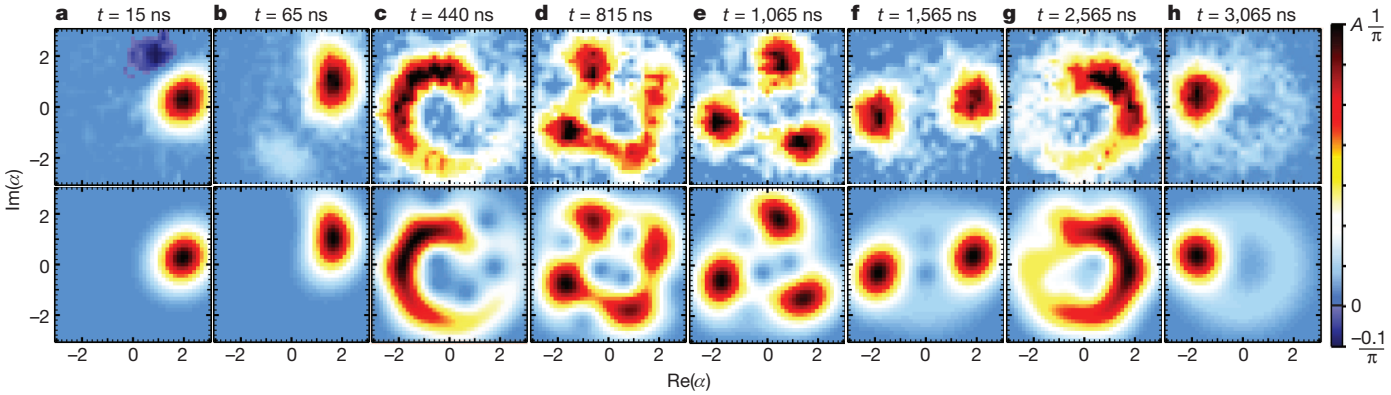


Figure 3 | Time evolution of Q_0 for a coherent state in the nonlinear cavity. Shown are experiment (upper row) and theory (lower row) for a coherent state $\beta = 2$; the time t for the frames shown in **a**–**h** is given above each panel. We measure Q_0 at 441 different analysis displacements α. The resolution of the pictures was doubled by interpolation. Initially the phase of the state spreads rapidly, **a**–**c**, leading to a complete phase collapse after a characteristic time $T_{\mathrm{col}} = \dfrac{\pi}{2\sqrt{\bar{n}}K} = 385$ ns. For short times, the Kerr interaction leads to a quadrature squeezed state along the Re(α) axis, which can be seen in **b**. After the complete phase collapse, structure emerges again, **d**–**h**, at times which are integer submultiples of the complete revival time $T_{\mathrm{rev}} = \dfrac{2\pi}{K}$. At these times, one obtains coherent state superpositions which are multi-component cat states, up to a maximum number of resolvable components set by the average photon number in the initial coherent state displacement. The colour scale for Q_0 of each pair of plots is individually rescaled, with the scaling parameter given for **a**–**h** respectively by $A = 1, 0.93, 0.32, 0.24, 0.3, 0.46, 0.24, 0.89$. The negative amplitudes in plot **a** are due to a 10% excited state population of the qubit which evolves in a different rotating frame (see Supplementary Information). This excited state population is not visible in the other frames as it disperses quickly and vanishes in the large positive amplitudes.

the measured state ρ_{m}, compared to an ideal n-component cat state $|\Psi_{\mathrm{id}}\rangle$, consisting of coherent states with amplitude $|\beta| = 2e^{-\kappa t/2}$, is $F_2 = 0.71$, $F_3 = 0.70$, $F_4 = 0.71$ for the two-, three-, four-component cat states, respectively. The Wigner functions show clear interference fringes, which demonstrates that the evolution is indeed coherent and well described by the wavefunction given in equation (2). The main reduction in the fidelity is due to the spurious excited state population of the qubit (see Supplementary Information) and the decay of the resonator state. The decay of the resonator state is also responsible for the asymmetry in the interference fringes of the Wigner function—for example, the maximum of the interference fringes for the two-component cat states is shifted to the left in both theory and experiment.

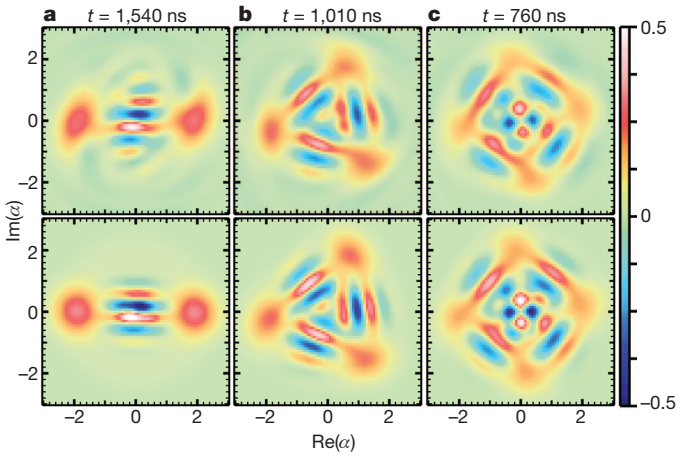


Figure 4 | Wigner function of the multi-component cat states emerging during the Kerr interaction. The top row shows the Wigner functions, reconstructed by cavity state tomography, of a coherent state subject to a Kerr interaction for a time t. The lower row shows the theoretically expected Wigner functions for the same interaction time obtained by a simulation including the decay of the cavity. The Wigner functions are reconstructed from measurements of the quasi probability distributions $Q_n(\alpha)$ for $n = 0$–7. Each $Q_n(\alpha)$ was measured at the same displacements as the corresponding data in Fig. 3. Comparing the experimentally obtained state to an ideal cat state, we find a fidelity $F_2 = 0.71$, $F_3 = 0.70$, $F_4 = 0.71$ for the two-, three- and four-component cats respectively. The experimental data show excellent correspondence with theoretical predictions, including the interference fringes and regions of negative quasi-probability distribution, confirming that highly non-classical states are produced by the Kerr evolution.

We have shown that we can engineer strong photon–photon interactions in a cavity, entering the single-photon Kerr regime where $K \gg \kappa$. We are able to observe the collapse and revival of a coherent state due to the intensity-dependent dispersion between Fock states in the cavity. This opens the possibility of using such a Kerr medium for error correction schemes where a nonlinear cavity is used to realize the necessary components[9]. The good agreement between the theory and the experiment demonstrates the accurate understanding of this system. It also confirms our ability to predict higher-order couplings, which is a necessary ingredient for understanding the behaviour of large circuit QED systems. Furthermore, we have measured the evolution of a coherent state in a Kerr medium at the single-photon level, and shown a new experimental way for creating and measuring multi-component Schrödinger cat states. This demonstrates the ability to create, manipulate and visualize coherent states in a larger Hilbert space, and opens up new directions for continuous variable quantum computation[30].

Received 17 October 2012; accepted 11 January 2013.

1. Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431,** 162–167 (2004).
2. Haroche, S. & Raimond, J. M. *Exploring the Quantum: Atoms, Cavities, And Photons* (Oxford Univ. Press, 2006).
3. Hofheinz, M. *et al.* Synthesizing arbitrary quantum states in a superconducting resonator. *Nature* **459,** 546–549 (2009).
4. Paik, H. *et al.* Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107,** 240501 (2011).
5. Yurke, B. & Stoler, D. The dynamic generation of Schrodinger cats and their detection. *Physica B+C* **151,** 298–301 (1988).
6. Braunstein, S. L. Quantum information with continuous variables. *Rev. Mod. Phys.* **77,** 513–577 (2005).
7. Grangier, P., Levenson, J. & Poizat, J. Quantum non-demolition measurements in optics. *Nature* **396,** 537–542 (1998).
8. Peyronel, T. *et al.* Quantum nonlinear optics with single photons enabled by strongly interacting atoms. *Nature* **488,** 57–60 (2012).
9. Kerckhoff, J., Nurdin, H., Pavlichin, D. S. & Mabuchi, H. Designing quantum memories with embedded control: photonic circuits for autonomous quantum error correction. *Phys. Rev. Lett.* **105,** 040502 (2010).
10. Milburn, G. Quantum optical Fredkin gate. *Phys. Rev. Lett.* **62,** 2124–2127 (1989).
11. Saleh, B. E. A. & Teich, M. C. *Fundamentals of Photonics* (Wiley Series in Pure and Applied Optics, 1991).

©2013 Macmillan Publishers Limited. All rights reserved

12. Slusher, R. E., Hollberg, L., Yurke, B., Mertz, J. & Valley, J. Observation of squeezed states generated by four-wave mixing in an optical cavity. *Phys. Rev. Lett.* **55,** 2409–2412 (1985).
13. Franken, P., Hill, A., Peters, C. & Weinreich, G. Generation of optical harmonics. *Phys. Rev. Lett.* **7,** 118–119 (1961).
14. Fisher, R. A., Kelley, P. L. & Gustafson, T. K. Subpicosecond pulse generation using the optical Kerr effect. *Appl. Phys. Lett.* **14,** 140–143 (1969).
15. Nigg, S. E. *et al.* Black-box superconducting circuit quantization. *Phys. Rev. Lett.* **108,** 240502 (2012).
16. Bourassa, J., Beaudoin, F., Gambetta, J. M. & Blais, A. Josephson-junction-embedded transmissionline resonators: from Kerr medium to in-line transmon. *Phys. Rev. A* **86,** 013814 (2012).
17. Bergeal, N. *et al.* Phase-preserving amplification near the quantum limit with a Josephson ring modulator. *Nature* **465,** 64–68 (2010).
18. Yurke, B. *et al.* Observation of 4.2-K equilibrium-noise squeezing via a Josephson-parametric amplifier. *Phys. Rev. Lett.* **60,** 764–767 (1988).
19. Yin, Y. *et al.* Dynamic quantum Kerr effect in circuit quantum electrodynamics. *Phys. Rev. A* **85,** 023826 (2012).
20. Hoffman, A. *et al.* Dispersive photon blockade in a superconducting circuit. *Phys. Rev. Lett.* **107,** 053602 (2011).
21. Shalibo, Y. *et al.* Direct Wigner tomography of a superconducting anharmonic oscillator. Preprint at http://arXiv.org/abs/1208.2441 (2012).
22. Meekhof, D. M., Monroe, C., King, B. E., Itano, W. M. & Wineland, D. J. Generation of nonclassical motional states of a trapped atom. *Phys. Rev. Lett.* **76,** 1796–1799 (1996).
23. Greiner, M., Mandel, O., Hänsch, T. W. & Bloch, I. Collapse and revival of the matter wave field of a Bose-Einstein condensate. *Nature* **419,** 51–54 (2002).
24. Johnson, B. R. *et al.* Quantum non-demolition detection of single microwave photons in a circuit. *Nature Phys.* **6,** 663–667 (2010).
25. Solano, E. Selective interactions in trapped ions: state reconstruction and quantum logic. *Phys. Rev. A* **71,** 013813 (2005).
26. Milburn, G. Quantum and classical Liouville dynamics of the anharmonic-oscillator. *Phys. Rev. A* **33,** 674–685 (1986).
27. Deléglise, S. *et al.* Reconstruction of non-classical cavity field states with snapshots of their decoherence. *Nature* **455,** 510–514 (2008).
28. Eichler, C., Bozyigit, D. & Wallraff, A. Characterizing quantum microwave radiation and its entanglement with superconducting qubits using linear detectors. *Phys. Rev. A* **86,** 032106 (2012).
29. Leibfried, D. *et al.* Experimental determination of the motional quantum state of a trapped atom. *Phys. Rev. Lett.* **77,** 4281–4285 (1996).
30. Leghtas, Z. *et al.* Hardware-efficient autonomous quantum error correction. Preprint at http://arXiv.org/abs/1207.0679 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. H. Devoret, M. D. Reed, M. Hatridge and A. Sears for discussions. This research was supported by the National Science Foundation (NSF) (PHY-0969725), the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) through the Army Research Office (W911NF-09-1-0369), and the US Army Research Office (W911NF-09-1-0514). Use of facilities was supported by the Yale Institute for Nanoscience and Quantum Engineering (YINQE) and the NSF (MRSECDMR 1119826). S.M.G. acknowledges support from the NSF (DMR-1004406). M.M. and Z.L. acknowledge support from French Agence Nationale de la Recherche under the project EPOQ2 (ANR-09-JCJC-0070). S.E.N. acknowledges support from the Swiss NSF. E.G. acknowledges support from EPSRC (EP/I026231/1).

Author Contributions G.K. and B.V. performed the experiments. G.K., B.V., Z.L. and M.M. analysed the data. G.K., B.V. and L.F. fabricated the qubits. G.K., B.V., H.P. and S.E.N. designed the device and calculated the device parameters. G.K., E.G., S.M.G. and R.J.S. conceived the experiment and all authors co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.J.S. (robert.schoelkopf@yale.edu).

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature11915

Coherent state transfer between itinerant microwave fields and a mechanical oscillator

T. A. Palomaki[1,2], J. W. Harlow[1,2], J. D. Teufel[3], R. W. Simmonds[3] & K. W. Lehnert[1,2]

Macroscopic mechanical oscillators have been coaxed into a regime of quantum behaviour by direct refrigeration[1] or a combination of refrigeration and laser-like cooling[2,3]. This result supports the idea that mechanical oscillators may perform useful functions in the processing of quantum information with superconducting circuits[4–7], either by serving as a quantum memory for the ephemeral state of a microwave field or by providing a quantum interface between otherwise incompatible systems[8–14]. As yet, the transfer of an itinerant state or a propagating mode of a microwave field to and from a storage medium has not been demonstrated, owing to the inability to turn on and off the interaction between the microwave field and the medium sufficiently quickly. Here we demonstrate that the state of an itinerant microwave field can be coherently transferred into, stored in and retrieved from a mechanical oscillator with amplitudes at the single-quantum level. Crucially, the time to capture and to retrieve the microwave state is shorter than the quantum state lifetime of the mechanical oscillator. In this quantum regime, the mechanical oscillator can both store quantum information and enable its transfer between otherwise incompatible systems.

Mechanical oscillators are particularly appealing for storing or transducing quantum information encoded in microwave fields, because their fabrication is compatible with, and they are similar in size to, superconducting quantum circuits. Indeed, a high-frequency mechanical oscillator has been combined with a superconducting quantum bit (qubit), demonstrating the transfer of a qubit state to the oscillator[1]. Although that was an impressive demonstration, the few-nanosecond lifetime of the oscillator, which was much shorter than the lifetime of the qubit, suggests that lower-frequency oscillators with much longer lifetimes may form superior quantum memories and transducers. This is particularly true for storing the information in an itinerant mode of a microwave field, for which the characteristic time to acquire a new value is about 100 ns (refs 15–17). The penalty for working with a low-frequency oscillator is that the oscillator will not naturally be in its quantum ground state, but rather a hot thermal state. For example, an oscillator with a resonance frequency of $\Omega_m = 2\pi \times 10$ MHz in equilibrium with an environment at temperature $T_{env} = 25$ mK will contain a statistically fluctuating number of quanta with average value $n_{env} \equiv [\exp(\hbar\Omega_m/k_B T_{env}) - 1]^{-1} = 50$ quanta, where k_B is the Boltzmann constant and $\hbar$ is Planck's constant divided by 2π. Fortunately, embedding the mechanical oscillator in a high-frequency resonant circuit creates an interaction between the mechanical oscillator and itinerant microwave fields that can be used cool the oscillator to its ground state[2].

Here we show that this same interaction can exchange the state of an itinerant microwave field with the state of the macroscopic mechanical oscillator. This exchange simultaneously transfers the information from travelling microwave fields in a transmission line to the mechanical oscillator, while converting the state of the oscillator to itinerant microwaves. Because the microwave fields are in a pure state, this prepares the oscillator in a low-entropy state with an average thermal noise below the single-quantum level. We demonstrate that this low-entropy state is preserved for a characteristic time of 90 μs, which is about 190 times longer than the minimum transfer time.

Our mechanical oscillator is the fundamental drumhead mode of a thin (100 nm), 15-μm-diameter superconducting aluminium membrane, which forms the upper plate of a 50-nm vacuum gap capacitor[18] (Fig. 1a). This arrangement forms a parallel-plate capacitor that, together with a spiral inductor, creates a microwave resonant circuit with resonant frequency $\omega_c \approx 2\pi \times 7.5$ GHz. We couple energy to and from the circuit by means of a nearby transmission line, and determine the device parameters from spectroscopic measurements (similar to ref. 2) in a dilution refrigerator with a base temperature of 15 mK. The tension and diameter of the membrane result in a fundamental frequency $\Omega_m = 2\pi \times 10.5$ MHz, with a linewidth $\gamma_m = 2\pi \times 35$ Hz. This frequency and the oscillator mass, $m = 48$ pg, imply a zero-point motion $x_{zp} = 4.1$ fm.

The key to our state transfer measurements is the ability to switch the coupling between the mechanical oscillator and the resonant circuit rapidly on and off. That the two systems are coupled can be understood from the fact that the motion of the oscillator changes

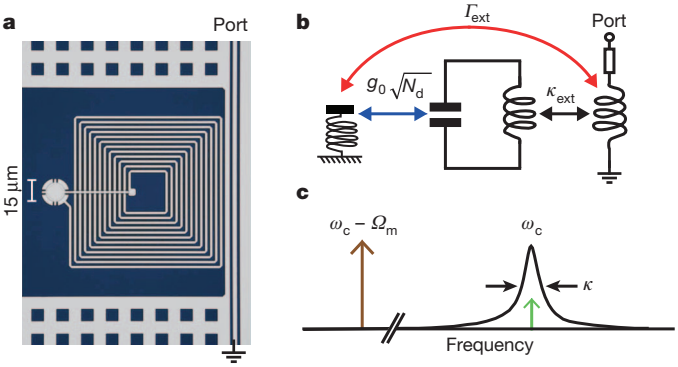


Figure 1 | Schematic description of the experiment. a, False-colour image of the device, which comprises a sputtered aluminium circuit (light grey) on a sapphire substrate (blue). A coplanar-waveguide transmission line visible at the right edge of the image couples itinerant microwave fields to the resonant circuit via mutual inductance with the spiral inductor of the resonator. One end of the transmission line is terminated in a short circuit (indicated by the earth symbol), which reflects microwave fields back to the input port. The mechanical oscillator is the 15-μm disk visible at the left edge of the figure. **b**, The diagram represents the exchange of energy between itinerant microwave fields and mechanical motion. In the presence of a transfer field with N_d photons, the coupling rate between the resonator and the oscillator fields is $g_0\sqrt{N_d}$. The resonator energy enters from and departs to the transmission line at a rate κ_{ext}. The transfer rate between itinerant microwave fields and mechanical motion is Γ_{ext}. The preparation and transfer field are applied and measured through the port. **c**, A frequency-domain representation of the experiment shows that the preparation field (green) and the transfer field (brown) are widely separated ($\Omega_m \gg \kappa$) compared with the linewidth of the circuit's response (black).

[1]JILA, National Institute of Standards and Technology and the University of Colorado, Boulder, Colorado 80309, USA. [2]Department of Physics, University of Colorado, Boulder, Colorado 80309, USA. [3]National Institute of Standards and Technology, Boulder, Colorado 80305, USA.

©2013 Macmillan Publishers Limited. All rights reserved

the circuit's resonance frequency, ω_c. The coupling can be described by an interaction Hamiltonian $H_{int} = \hbar G\hat{x}\hat{a}^\dagger\hat{a}$, where $G = d\omega_c/dx$, $\hat{a}$ and $\hat{a}^\dagger$ respectively correspond to the annihilation and creation operators for microwave photons, $\hat{x} = x_{zp}\left(\hat{b}^\dagger + \hat{b}\right)$ is the position operator of the mechanical oscillator, and $\hat{b}$ and $\hat{b}^\dagger$ respectively are the annihilation and creation operators for mechanical phonons. In the presence of a strong microwave excitation at frequency $\omega_d = \omega_c - \Omega_m$, we can express $\hat{a}$ in a frame co-rotating at ω_d as $\hat{a} = \alpha + \hat{d}$. The interaction between the two systems is greatly enhanced and takes the approximate form[14]

$$H_{int} = \hbar g_0\sqrt{N_d(t)}\left(\hat{d}\hat{b}^\dagger + \hat{b}\hat{d}^\dagger\right) \quad (1)$$

where $N_d(t) = |\alpha|^2$ is the strength of the excitation at ω_d expressed as the number of photons in the resonator, and $g_0 = Gx_{zp} = 2\pi \times 200$ Hz. The interaction Hamiltonian now resembles that of a beam splitter and as such it coherently exchanges energy between microwave fields at ω_c and mechanical phonons, as highlighted by the blue arrow in Fig. 1b. Crucially, the rate of that exchange—the transmission of the beam splitter—is controlled by $N_d(t)$, a quantity that we can change. In addition to the dynamics contained within equation (1), energy is also coupled between the resonant circuit and itinerant modes in the transmission line at a rate $\kappa_{ext} = 2\pi \times 275$ kHz. The resonator's linewidth is $\kappa = \kappa_{in} + \kappa_{ext}$, where κ_{in} is the internal loss rate, which depends on the energy in the resonant circuit[19]. For the range of microwave powers used here, we find that κ is between $2\pi \times 320$ kHz and $2\pi \times 360$ kHz. Equation (1), which permits the rapid and coherent exchange between electrical and mechanical energy, is an approximation which assumes that κ is small enough to satisfy $\kappa \ll \Omega_m$ (the resolved sideband limit).

For linewidths satisfying the resolved sideband limit, the microwave fields incident on the circuit can be divided into two types of field on the basis of their frequencies (Fig. 1c). Fields with frequencies near ω_c (preparation fields) prepare the itinerant mode in a particular state, which can be transferred into the mechanical oscillator. Fields with frequencies near $\omega_c - \Omega_m$ (transfer fields) transfer states between the mechanical oscillator and the microwave circuit. In the resolved sideband limit (that is, for $\kappa \ll \Omega_m$), these fields can remain spectrally distinct even when their strengths are varied more rapidly than κ. The transfer field's strength (Fig. 1b) determines whether itinerant fields in a nearby transmission line or states of the resonator are transferred. Because the resonator's state evolves at rate κ, if $2g_0\sqrt{N_d} < \kappa$ then the resonator never completely contains the former or future state of the mechanical oscillator and the transfer process is between itinerant microwave fields and the oscillator[20]. If, instead, the coupling rate is increased such that $2g_0\sqrt{N_d} > \kappa$ (the strong-coupling regime), the state of the resonator is swapped with that of the oscillator, setting the bandwidth limit of the transfer process for itinerant fields and the oscillator[21–23] (Supplementary Information).

The phenomena of cooling, state measurement and optomechanically induced transparency can be understood as steady-state limits of the transfer process in one regime or the other. For example, if the preparation field is unexcited in its vacuum state, continuous application of the transfer field exchanges the thermal state of the mechanical oscillator for the ground state of the microwave field, cooling the oscillator[24–26]. Microwave fields emitted by the circuit carry away the thermal state, which can then be measured. If, instead, a transfer field and a preparation field are both applied continuously, they produce an interference effect in the response of the circuit to the preparation field (optomechanically induced transparency). All of these phenomena are consequences of the classical equations of motion for the coupled system; however, if the transfer rate between phonons and itinerant photons

$$\Gamma_{ext} = \frac{4g_0^2 N_d}{\kappa}\frac{\kappa_{ext}}{\kappa} \quad (2)$$

exceeds the rate at which a single phonon is exchanged with the oscillator's environment, the transfer process is capable of exchanging single mechanical phonons and microwave photons. The factor κ_{ext}/κ in equation (2) distinguishes the transfer rate to itinerant photons from the total transfer rate to all channels, $\Gamma = 4g_0^2 N_d/\kappa$. A single phonon enters from the environment at a rate given by $\sim n_{env}\gamma_m$. The ratio of the total transfer rate to the environmental decoherence rate takes the form of a cooperativity parameter, $\tilde{C} = 4g_0^2 N_d/\kappa n_{env}\gamma_m$. For $\tilde{C} > 1$, the system is in the quantum regime. The device we study is well suited to demonstrating state transfer because its linewidth satisfies the resolved sideband limit and it is capable of achieving $\tilde{C} \gg 1$.

Efficiently transferring a state encoded in a pulse requires that the transfer process selects only the temporal envelope, or mode, of that pulse. From the equations of motion, it is evident that by controlling the envelope of the transfer field we should be able to choose the mode of the itinerant microwave fields transferred into or out of the macroscopic oscillator[27–30]. For example, a constant transfer field maps the state of the mechanical oscillator into an itinerant microwave mode whose amplitude decays exponentially with rate $\Gamma/2$. Similarly, a constant transfer field captures an itinerant mode whose amplitude grows exponentially at rate $\Gamma/2$.

In Fig. 2, we verify our ability to select particular modes and determine the state transfer efficiency. First we apply both a constant transfer field and a preparation field, $V_{in}(t) = V_{prep}\exp\left(\Gamma_{prep}t/2\right)\cos(\omega_c t + \phi)\Theta(-t)$ (where $\Theta(t)$ is the unit step function), with energy $E_{prep} = \int_{-\infty}^{0} V_{in}^2(t)\,dt$ and $\Gamma_{prep} = 2\pi \times 6.4$ kHz, capturing some fraction of the preparation field in the oscillator. We store the captured fraction by turning off the transfer field at $t = 0$ for a time $\tau_{del} = 25$ μs. To retrieve the state from the mechanical oscillator, we apply the transfer field again (Fig. 2a). The voltage wave generated in the transmission line by the circuit, $V_{out}(t)$, is amplified using a sensitive microwave circuit and mixed down to a frequency $\omega_{out}/2\pi \approx 1$ MHz, at which it can be readily digitized. In Fig. 2b, we plot the coherent portion of $V_{out}(t)$ as $\langle V_{out}(t)/V_{prep}\rangle$, inferred from the digitized voltage by averaging 600 repetitions of the protocol. The preparation field can bypass the oscillator and be measured at $t < \tau_{del}$; be captured by the oscillator, stored, retrieved and then measured at $t > \tau_{del}$; or be dissipated in the resonator or oscillator. When $\Gamma \approx \Gamma_{prep}$, the transferred mode is well matched to the preparation pulse, as seen from the nearly complete extinction of the bypassed component. In contrast, the transferred mode is poorly matched to the preparation pulse when $\Gamma \ll \Gamma_{prep}$ or $\Gamma \gg \Gamma_{prep}$. In Fig. 2c, we plot the measured coherent energy, E, normalized by the preparation energy, for both the bypassed field, $(1/E_{prep})\int_{-\infty}^{\tau_{del}} \langle V_{out}(t)\rangle^2\,dt$, and the retrieved field, $(1/E_{prep})\int_{\tau_{del}}^{\infty} \langle V_{out}(t)\rangle^2\,dt$. We find that when the total transfer rate, Γ, is equal to Γ_{prep}, 65% of the energy is captured, stored and retrieved. Because the protocol comprises two state transfers (a capture and a retrieval) we infer an efficiency of $\eta_{st} = 0.81 \pm 0.02 = \sqrt{0.65}$ for a single transfer. The equations of motion for the system imply that the maximum possible transfer efficiency is $\eta_{st} = \kappa_{ext}/\kappa = 0.80 \pm 0.01$. Indeed, we find excellent agreement between Fig. 2 and the expected behaviour with all device parameters independently determined.

Figure 2 demonstrates a transfer process capable of exchanging energy coherently; to determine its potential for quantum applications, we measure the noise properties of the transfer process through the statistics of an ensemble of measurements. We calibrate the noise in units of quanta by first preparing the oscillator in a known thermal state. Namely, we set $\tau_{del} = 75$ ms $\gg 1/\gamma_m$, allowing the oscillator to equilibrate with a known environment. When the transfer field is turned on, the state of the mechanical oscillator is mapped to a mode of the microwave field that decays exponentially at a rate $\Gamma/2$. The phase and amplitude of $V_{out}(t)$ after turning on the transfer field encode the state of the mechanical oscillator just before the field was energized. Using our knowledge of Γ and ω_{out}, we can extract the amplitude and phase of the microwave temporal mode from each realization of the measurement, and thus infer the state of the oscillator

©2013 Macmillan Publishers Limited. All rights reserved

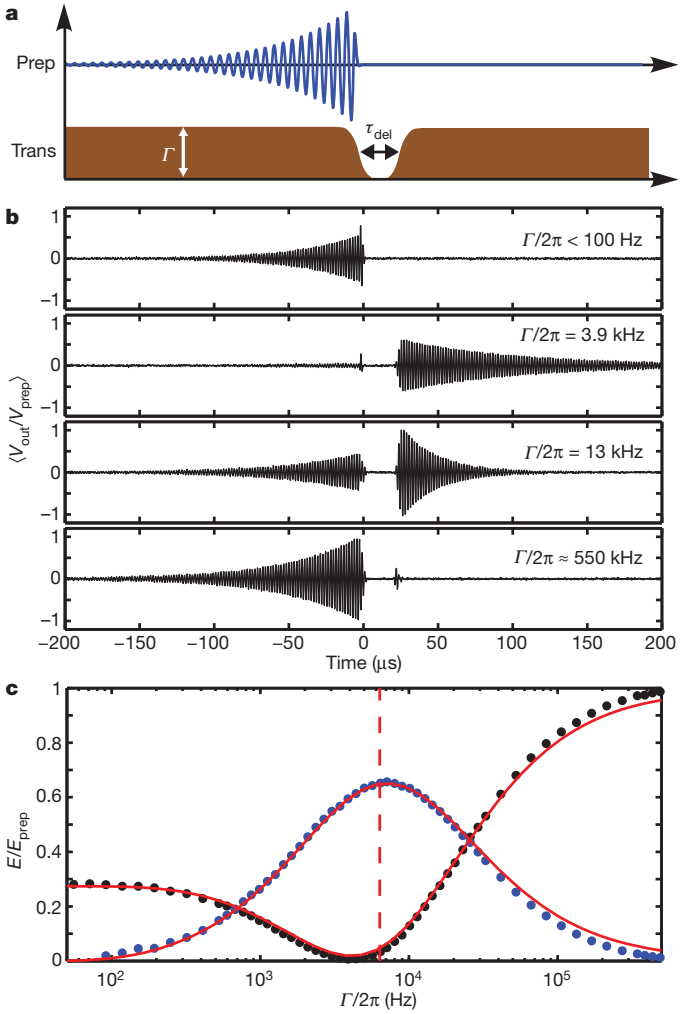


Figure 2 | Coherent state transfer. **a**, The diagram depicts the protocol for capture, storage and retrieval for an itinerant microwave field. The upper trace illustrates the preparation field, which has an exponential envelope rising at rate $\Gamma_{\mathrm{prep}}/2 = 2\pi \times 3.2$ kHz. **b**, Only when $\Gamma \approx \Gamma_{\mathrm{prep}}$ do we find substantial extinction of the preparation pulse and efficient capture, storage and retrieval. (The respective transfer rates for exact extinction and maximum efficiency coincide only if $\kappa_{\mathrm{in}} = 0$.) **c**, The coherent microwave energy, E, captured, stored and retrieved from the mechanical oscillator (blue points) is plotted as a function of the total transfer rate, Γ, where the y axis has been normalized to the total energy in the preparation field. At the point of maximum efficiency, we find that 65% of the coherent energy is captured, stored and retrieved. The black points show the fraction of the preparation field that bypasses the oscillator before being measured. The solid red lines are the expected results with all parameters determined separately, and the dashed line marks $\Gamma = \Gamma_{\mathrm{prep}}$.

just before it was transferred. Rather than work with amplitude and phase variables, we prefer to describe the state of the oscillator in terms of quadrature amplitudes, $\hat{X}_1 = \left(\hat{b}e^{i\Omega_m t} + \hat{b}^\dagger e^{-i\Omega_m t}\right)/2$ and $\hat{X}_2 = \left(\hat{b}e^{i\Omega_m t} - \hat{b}^\dagger e^{-i\Omega_m t}\right)/2i$, which have the virtue of being canonically conjugate to each other. We refer to the inferred values of $\hat{X}_1$ and $\hat{X}_2$ in any single measurement as X_1 and X_2, respectively.

This inference is not perfect because noise with both quantum and technical origins is added by the state transfer and subsequent measurement of the microwave field. We independently determine the variance of this noise and denote it $1/\eta = 1/\eta_m\eta_{st}$, where $\eta_m \leq 1$ is the effective quantum efficiency for the measurement of the microwave state. For a mechanical oscillator in equilibrium with an environment at temperature T_{env}, we expect the measurements of X_1 and X_2 to follow a two-dimensional Gaussian distribution whose variance depends on the environment's temperature as $\langle X_1^2 + X_2^2 \rangle \propto n_{\mathrm{env}} + 1/\eta$, where the brackets denote an ensemble average. Using a single normalizing factor, we can then express X_1 and X_2 in 'quantum units' such that $\langle X_1^2 + X_2^2 \rangle = n_{\mathrm{env}} + 1/\eta$ (Supplementary Information). In Fig. 3a, we plot 2,000 measurements of the oscillator's state with $T_{\mathrm{env}} = 25$ mK and find a distribution consistent with a thermal state.

Having calibrated the measurements into quantum units, we now demonstrate that the transfer process can prepare states of the oscillator with a particular amplitude and phase, but with less than one quantum of noise. We apply the protocol of Fig. 2a with $\tau_{\mathrm{del}} = 25$ μs and $\Gamma = 2\pi \times 7$ kHz$\approx 4n_{\mathrm{env}}\gamma_m$. We choose a preparation pulse containing 35 quanta and one of four different phases, ϕ, chosen from the set $\{\pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$. Figure 3b shows the measurements of the state of the oscillator for 8,000 repetitions of the protocol, with 2,000 per phase. The data show that the phase of the preparation pulse is faithfully recovered after storage. Furthermore, the total variances in Fig. 3b are 13 times smaller than in Fig. 3a, signifying a lower entropy and thermal energy. We can distinguish an energy associated with the coherent component as $N_{\mathrm{coh}} = \langle X_1 \rangle^2 + \langle X_2 \rangle^2$ and can identify the thermal, or noise, component as $N_{\mathrm{th}} = \langle (X_1 - \langle X_1 \rangle)^2 \rangle + \langle (X_2 - \langle X_2 \rangle)^2 \rangle - 1/\eta$. From Fig. 3b, we then find a measurement efficiency $1/\eta = 3.1 \pm 0.1$ (Supplementary Information), a coherent component $N_{\mathrm{coh}} = 28.7 \pm 0.8$ and a thermal component $N_{\mathrm{th}} = 0.9 \pm 0.1$, where the stated uncertainty is the standard deviation of the four measurements. Although the average values, $\langle X_1 \rangle$ and $\langle X_2 \rangle$, can be determined from the classical equations of motion, we cannot account for the distribution of X_1 and X_2 without including the quantum noise of measurement (Fig. 3b). In previous measurements where the oscillator was simply cooled with a steady transfer field at $\Gamma = 2\pi \times 7$ kHz, we found a thermal occupancy of 0.5 quanta[2]. Because the oscillator should acquire 0.3 quanta during the 25-μs storage period, we would expect the oscillator to have a thermal occupancy of 0.8 quanta just before the retrieval transfer. These results show that by state transfer we can prepare the oscillator in a particular state with less than one noise quantum and measure that state with sensitivity near the quantum limit.

Having demonstrated the ability to prepare the mechanical oscillator in a low-entropy displaced state, we can estimate the quantum state lifetime by observing the evolution of the state back into a thermal state in equilibrium with its environment. Figure 3c shows the evolution of N_{coh} and N_{th} as a function of τ_{del}, where we have chosen a preparation field such that $N_{\mathrm{coh}}(0) \approx n_{\mathrm{env}} = 50$. As expected, the coherent component exponentially decays while the entropy returns to the mechanical oscillator, both with rate γ_m. We find excellent agreement between the data and the predicted evolution. Thus, we can estimate the quantum state lifetime as the characteristic time to add a single phonon from the environment, $1/n_{\mathrm{env}}\gamma_m = 90$ μs.

For quantum applications, we envision transferring states with energy around one quantum. To test our transfer process in this regime, and to determine the accuracy of our state preparation protocol, we repeat the procedure of Fig. 2b, but reduce the preparation field amplitude such that the incident pulse contains approximately one quantum of energy, and again set $\tau_{\mathrm{del}} = 25$ μs. By separately measuring the preparation pulse and using our knowledge of the transfer efficiency, we anticipate the phase and amplitude that we will retrieve from the oscillator. We compare this expected value to the measured values of $\langle X_1 \rangle$ and $\langle X_2 \rangle$ averaged from 2,000 repetitions of the protocol (Fig. 3d). We find that the deviation of the measured values from the expected values is consistent with purely statistical error and is below 0.1 quanta for this preparation field, indicating that our state preparation protocol is accurate at this level.

Looking to the future, the ability to store microwave states in a low-loss mechanical oscillator provides an attractive tool in the growing field of quantum technology. Although our measurements here have dealt with classical states with energies on the level of single quanta, we can readily imagine incorporating a superconducting qubit to generate itinerant Fock states[15,17] that could be stored in the mechanical oscillator[10,14]. Our measurements also represent a step towards the more ambitious

©2013 Macmillan Publishers Limited. All rights reserved

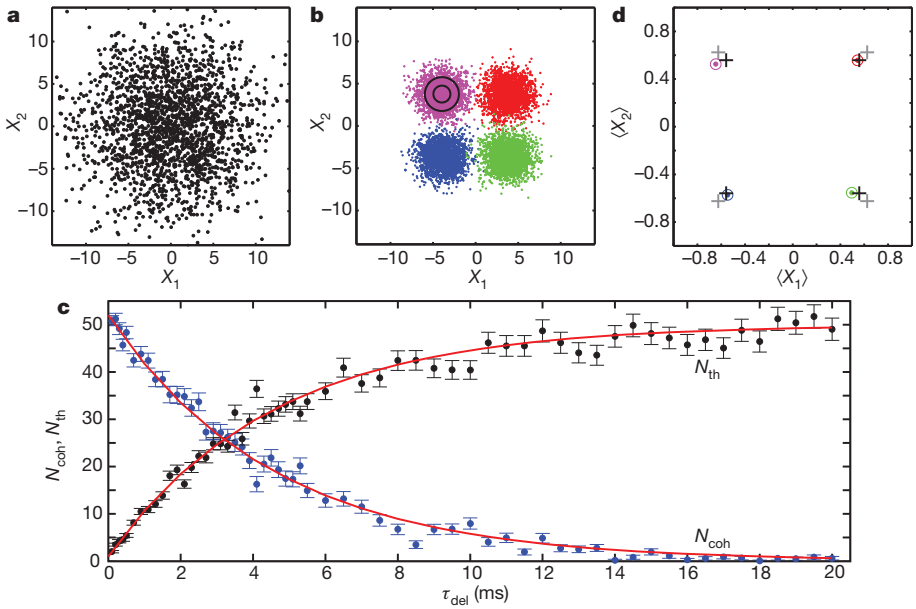


Figure 3 | A mechanical oscillator as a phase-coherent memory. **a**, $V_{out}(t)$ is measured after the application of a transfer field and yields a single point in quadrature phase space that represents the best estimate of the state of the mechanical oscillator. The points are 2,000 independent measurements of the state of the mechanical oscillator when in equilibrium at $T_{env} = 25$ mK. **b**, The phase of the preparation pulse in the protocol of Fig. 2a is rotated in $\pi/2$ increments with $\tau_{del} = 25$ μs and $\Gamma = 2\pi \times 7$ kHz. We repeat the protocol 8,000 times, 2,000 times per phase. Each measurement yields a single point in phase space that has been coloured to correspond to the phase of the preparation field. The radii of the circles on the magenta data are given by $\sqrt{N_{th} + 1/\eta}$, where the outer circle is the experimentally determined standard deviation and the inner circle is the quantum mechanical minimum, for which $N_{th} = 0$ and $\eta = 1$. **c**, Occupancy of the oscillator as a function of τ_{del} is decomposed into a thermal component (black) and a coherent component (blue), with 500 measurements for each delay. Error bars, s.e.m. The lines are the expected evolution of N_{th} and N_{coh} as predicted from the equations of motion and $\gamma_m = 2\pi \times 35$ Hz, which is determined separately from spectroscopic measurements. **d**, To demonstrate coherent state transfer at the level of a single quantum, we repeat the protocol but use preparation fields with coherent amplitudes six times smaller than in **b**. The points show the coherent amplitude and phase of the retrieved state, estimated by averaging the quadrature values of 2,000 measurements for each preparation phase. The circles enclosing the points indicate the standard error of the mean. The black crosses indicate the results expected from the state transfer efficiency determined in Fig. 2c and a separate measurement of the preparation pulse. The lighter crosses show what we would have expected to measure if the dominant source of inefficiency had been absent, that is, for $\kappa_{in} = 0$.

goal of transferring states between photons at microwave and optical frequencies[8–12] and using mechanical oscillators to generate entanglement[13]. As mechanical oscillators continue to develop in the quantum regime, new applications will emerge that exploit this quantum behaviour in the fields of fundamental physics, quantum information and precise force sensing. Our measurements demonstrate a novel means of controlling a mechanical oscillator, paving the way for many of these applications to become reality.

Received 14 June 2012; accepted 11 January 2013.

1. O'Connell, A. D. *et al.* Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464,** 697–703 (2010).
2. Teufel, J. D. *et al.* Sideband cooling of micromechanical motion to the quantum ground state. *Nature* **475,** 359–363 (2011).
3. Chan, J. *et al.* Laser cooling of a nanomechanical oscillator into its quantum ground state. *Nature* **478,** 89–92 (2011).
4. Hertzberg, J. B. *et al.* Back-action-evading measurements of nanomechanical motion. *Nature Phys.* **6,** 213–217 (2010).
5. Sillanpää, M. A., Sarkar, J., Sulkko, J., Muhonen, J. & Hakonen, P. J. Accessing nanomechanical resonators via a fast microwave circuit. *Appl. Phys. Lett.* **95,** 011909 (2009).
6. LaHaye, M. D., Suh, J., Echternach, P. M., Schwab, K. C. & Roukes, M. L. Nanomechanical measurements of a superconducting qubit. *Nature* **459,** 960–964 (2009).
7. Regal, C. A., Teufel, J. D. & Lehnert, K. W. Measuring nanomechanical motion with a microwave cavity interferometer. *Nature Phys.* **4,** 555–560 (2008).
8. Wang, Y. D. & Clerk, A. A. Using interference for high fidelity quantum state transfer in optomechanics. *Phys. Rev. Lett.* **108,** 153603 (2012).
9. Regal, C. A. & Lehnert, K. W. From cavity electromechanics to cavity optomechanics. *J. Phys. Conf. Ser.* **264,** 012025 (2011).
10. McGee, S. A., Meiser D., Regal, C. A., Lehnert, K. W. & Holland, M. J. Mechanical resonators for storage and transfer of electrical and optical quantum states. *Phys. Rev. A.* (submitted).
11. Tian, L. Adiabatic state conversion and pulse transmission in optomechanical systems. *Phys. Rev. Lett.* **108,** 153604 (2012).
12. Safavi-Naeini, A. H. & Painter, O. Proposal for an optomechanical traveling wave phonon-photon translator. *N. J. Phys.* **13,** 013017 (2011).
13. Hofer, S. G., Wieczorek, W., Aspelmeyer, M. & Hammerer, K. Quantum entanglement and teleportation in pulsed cavity optomechanics. *Phys. Rev. A* **84,** 052327 (2011).
14. Zhang, J., Peng, K. & Braunstein, S. L. Quantum-state transfer from light to macroscopic oscillators. *Phys. Rev. A* **68,** 013808 (2003).
15. Houck, A. A. *et al.* Generating single microwave photons in a circuit. *Nature* **449,** 328–331 (2007).
16. Mallet, F. *et al.* Quantum state tomography of an itinerant squeezed microwave field. *Phys. Rev. Lett.* **106,** 220502 (2011).
17. Eichler, C. *et al.* Experimental state tomography of itinerant single microwave photons. *Phys. Rev. Lett.* **106,** 259903 (2011).
18. Cicak, K. *et al.* Low-loss superconducting resonant circuits using vacuum-gap-based microwave components. *Appl. Phys. Lett.* **96,** 093502 (2010).
19. Gao, J. S. *et al.* Experimental evidence for a surface distribution of two-level systems in superconducting lithographed microwave resonators. *Appl. Phys. Lett.* **92,** 152505 (2008).
20. Fiore, V. *et al.* Storing optical information as a mechanical excitation in a silica optomechanical resonator. *Phys. Rev. Lett.* **107,** 133601 (2011).
21. Verhagen, E., Deleglise, S., Weis, S., Schliesser, A. & Kippenberg, T. J. Quantum-coherent coupling of a mechanical oscillator to an optical cavity mode. *Nature* **482,** 63–67 (2012).
22. Gröblacher, S., Hammerer, K., Vanner, M. R. & Aspelmeyer, M. Observation of strong coupling between a micromechanical resonator and an optical cavity field. *Nature* **460,** 724–727 (2009).
23. Teufel, J. D. *et al.* Circuit cavity electromechanics in the strong-coupling regime. *Nature* **471,** 204–208 (2011).
24. Teufel, J. D., Harlow, J. W., Regal, C. A. & Lehnert, K. W. Dynamical backaction of microwave fields on a nanomechanical oscillator. *Phys. Rev. Lett.* **101,** 197203 (2008).
25. Marquardt, F., Chen, J. P., Clerk, A. A. & Girvin, S. M. Quantum theory of cavity-assisted sideband cooling of mechanical motion. *Phys. Rev. Lett.* **99,** 093902 (2007).
26. Wilson-Rae, I., Nooshi, N., Zwerger, W. & Kippenberg, T. J. Theory of ground state cooling of a mechanical oscillator using dynamical backaction. *Phys. Rev. Lett.* **99,** 093901 (2007).
27. Novikova, I. *et al.* Optimal control of light pulse storage and retrieval. *Phys. Rev. Lett.* **98,** 243602 (2007).

©2013 Macmillan Publishers Limited. All rights reserved

28. Reim, K. F. *et al.* Single-photon-level quantum memory at room temperature. *Phys. Rev. Lett.* **107,** 053603 (2011).
29. Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* **78,** 3221–3224 (1997).
30. Boozer, A. D., Boca, A., Miller, R., Northup, T. E. & Kimble, H. J. Reversible state transfer between light and a single trapped atom. *Phys. Rev. Lett.* **98,** 193601 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported primarily by the DARPA QuASAR programme, with additional support from the US NSF Physics Frontier Center and NIST. We would like to thank J. Kerckhoff, M. Holland, C. Regal, J. Thompson and R. Andrews for discussions.

Author Contributions T.A.P. and K.W.L. designed the experiment, analysed the results and wrote the manuscript. T.A.P. performed the measurements. J.D.T. and R.W.S. designed and fabricated the device. J.W.H. provided experimental support. All authors commented on the results and manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.W.L. (konrad.lehnert@jila.colorado.edu).

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature11936

Mesoporous TiO_2 single crystals delivering enhanced mobility and optoelectronic device performance

Edward J. W. Crossland[1], Nakita Noel[1], Varun Sivaram[1], Tomas Leijtens[1], Jack A. Alexander-Webber[1] & Henry J. Snaith[1]

Mesoporous ceramics and semiconductors enable low-cost solar power, solar fuel, (photo)catalyst and electrical energy storage technologies[1]. State-of-the-art, printable high-surface-area electrodes are fabricated from thermally sintered pre-formed nanocrystals[2–5]. Mesoporosity provides the desired highly accessible surfaces but many applications also demand long-range electronic connectivity and structural coherence[6]. A mesoporous single-crystal (MSC) semiconductor can meet both criteria. Here we demonstrate a general synthetic method of growing semiconductor MSCs of anatase TiO_2 based on seeded nucleation and growth inside a mesoporous template immersed in a dilute reaction solution. We show that both isolated MSCs and ensembles incorporated into films have substantially higher conductivities and electron mobilities than does nanocrystalline TiO_2. Conventional nanocrystals, unlike MSCs, require in-film thermal sintering to reinforce electronic contact between particles, thus increasing fabrication cost, limiting the use of flexible substrates and precluding, for instance, multijunction solar cell processing. Using MSC films processed entirely below 150 °C, we have fabricated all-solid-state, low-temperature sensitized solar cells that have 7.3 per cent efficiency, the highest efficiency yet reported. These high-surface-area anatase single crystals will find application in many different technologies, and this generic synthetic strategy extends the possibility of mesoporous single-crystal growth to a range of functional ceramics and semiconductors.

Previous approaches to growing MSC semiconductors aimed at crystallizing the target material from thermal decomposition of precursor species infiltrated into the pores of a sacrificial template[7–17]. However, with the exception of surface-masked epitaxial growth from pre-grown crystal wafers[18], these approaches have achieved ambiguous success in realizing extended MSC semiconductors. Very often the resulting material is best described as a 'mesocrystal'—a porous assembly of nanoscale crystallites with common orientation but no single coherent atomic domain (the classical basis for a single crystal)[19,20]. The external shapes of such templated mesocrystals are spherical or otherwise irregular and do not express the underlying symmetry and facets of the crystal lattice, as is the case for non-porous single crystals. A classical route to large single crystals is growth in a dilute solution[21] and/or at reduced temperatures that lower the rate of nucleation and growth. At first this approach seems incompatible with a templating strategy because without a template volume that encompasses the entire reaction volume, only a vanishingly small number of crystals nucleate inside the template rather than in the bulk solution.

Here we demonstrate that the critical issue of confining growth to the guiding template in the dilute regime is solved by 'seeding' the template with microscopic nucleation sites for crystal growth, which then overwhelms homogeneous nucleation. The specific semiconductor crystals we report are structural inverses of a quasi-close-packed array of silica beads, with pore size tunable from 20 nm to 250 nm. By seeding this template, we achieve a near-unity yield of clearly faceted, high-surface-area single crystals of anatase TiO_2, whose external dimensions grow to orders of magnitude larger than their internal mesopores. There is no longer any need to overcome an energetic barrier to homogeneous nucleation, so the reaction temperature can be lowered significantly such that 'soft' materials like mesoporous block copolymers[22] become suitable host templates for fully crystalline functional ceramics. We propose that this general strategy is applicable to an enormous range of template morphologies and functional materials, from both solution and vapour-phase reactions, and has broad scope in functional ceramics and materials for energy generation and storage.

The anatase TiO_2 MSC growth is based on the synthesis of Yang *et al.*[23] via hydrothermal treatment of TiF_4 in the presence of hydrofluoric acid, which stabilizes the precursor and significantly reduces the nucleation rate[24,25]. A silica template with an external volume that is approximately 1% of the overall reaction volume is first seeded by pre-treatment in a solution of $TiCl_4$ before its immersion in the reaction vessel. The template is finally removed by selective etching in aqueous NaOH to recover the mesoporous TiO_2 crystal product, as illustrated in Fig. 1 (see Methods). Scanning electron microscope (SEM) characterization of the product (Fig. 1d–g) reveals faceted truncated bipyramidal crystals with external symmetry matching that of the homogeneously nucleated bulk crystals (Fig. 1c), whose mesoscale structure is a negative replica of the silica template. The crystal in Fig. 1d is particularly instructive, because it evidently grew from a seed located on or close to an external surface of the mesoporous template. In one direction the crystal growth is directed outwards into the solution and results in a compact, faceted TiO_2 crystal with a symmetry consistent with the anatase TiO_2 Wulff construction. In the opposite direction the growth front enters the templated volume such that the crystal grows around the guiding template (mesoscale detail shown in Fig. 1e). The single set of external crystal facets is strong evidence for the nucleation and growth mechanism of a single-crystal domain from a single nucleation site. Crystals grown entirely within the template pores (Fig. 1f and g) also have external symmetry matching the Wulff construction, now coupled with an internal mesopority whose accessibility is confirmed by dye adsorption–desorption surface area measurements of $150\,\mu m^2\,\mu m^{-3}$ or, equivalently, $70\,m^2\,g^{-1}$ for 20 nm pores in $2\,\mu m^3$ crystals. That is, the specific free surface of these MSCs is very similar to equivalent-sized nanoparticles (Dyesol DSL 18NR-T, 20 nm average particle size, $75\,m^2\,g^{-1}$).

Crucially, in the absence of seeding, the overwhelming majority of crystals nucleate and grow outside the template in the solution bulk, producing non-porous crystals of the type shown in Fig. 1c. The relative yield of non-seeded porous crystals is extremely low: only a careful inspection of the etched template product (a transparent solution containing no visible product) by SEM reveals a trace number of porous crystals that nucleate homogeneously inside the silica template. The total number of homogeneously nucleated crystals without seeding (N_0) follows from the average crystal volume and total TiO_2 product mass and is approximately 5×10^8. After seeding the template in 15 μM aqueous $TiCl_4$, however, the number of crystals increases to about 9×10^{10}—a nucleation rate roughly two orders of magnitude greater than the homogeneous rate. Seeded growth dominates homogeneous

[1]Clarendon Laboratory, University of Oxford Parks Road, Oxford, OX1 3PU, UK.

©2013 Macmillan Publishers Limited. All rights reserved

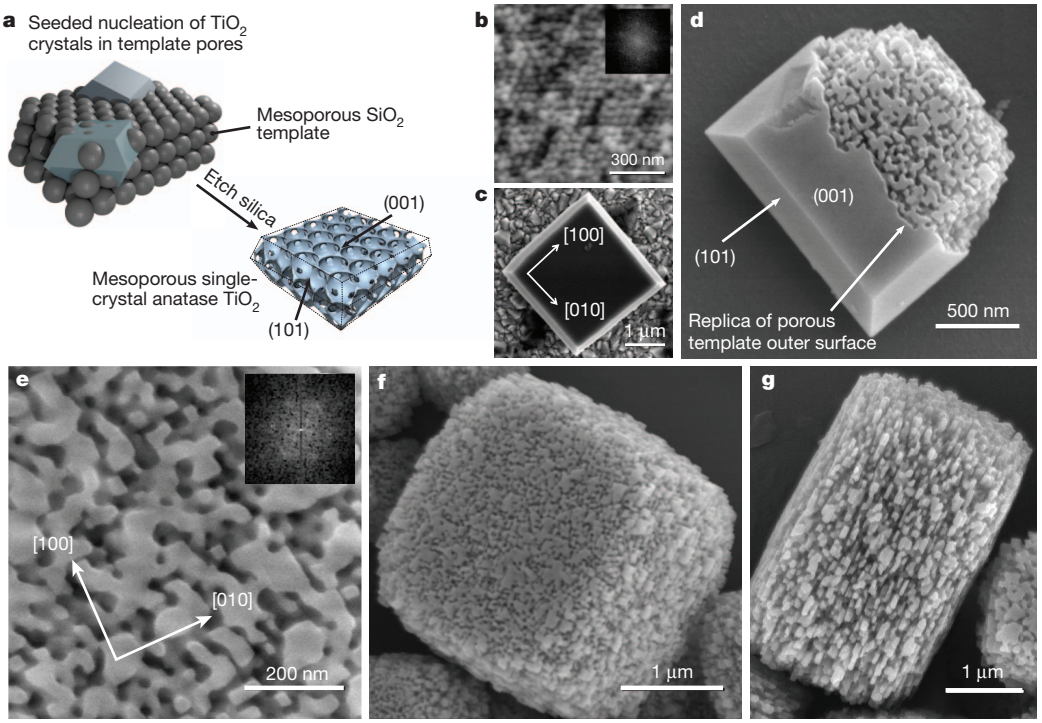


Figure 1 | MSC synthesis of TiO_2. **a**, Schematic of MSC nucleation and growth within a mesoporous template. **b**, Pristine silica template made up of quasi-close-packed silica beads (fast Fourier transform (FFT) inset showing sixfold symmetry with 49 ± 3 nm spacing). **c**, Non-porous truncated bipyramidal TiO_2 crystal homogeneously nucleated in a 20-mM TiF_4 solution at 210 °C. **d**, Template-nucleated variant of the crystal type shown in **c**, grown from a seed at or near the external template surface such that a non-porous volume coexists with a mesoporous region within a single faceted microcrystal. **e**, Replication of the mesoscale pore structure within the templated region (FFT inset, 47 ± 3 nm sixfold symmetry) with crystal lattice vectors implied from the particle symmetry overlaid (reaction conditions 170 °C, 40 mM TiF_4). **f** and **g**, Fully mesoporous TiO_2 crystals grown by seeded nucleation in the bulk of the silica template.

nucleation and the vast majority of TiO_2 crystals grow inside the template rather than in the bulk solution, giving a near-unity yield of mesoporous crystals.

We further probed the nature of the templated crystals at the atomic scale by X-ray diffraction (XRD) and electron diffraction, shown in Fig. 2. The X-ray diffraction pattern of an ensemble of the mesoporous product is compared in Fig. 2a to both a standard mesoporous anatase TiO_2 nanoparticle film (Dyesol DSL 18NR-T, 20 nm average particle size) and to non-templated (solid) anatase microcrystals. All patterns index to the anatase phase of titania. The apparent average domain size of the conventional nanoparticle film calculated from Scherrer broadening of the primary (101) peak is 20 nm. The peak width of the TiO_2 crystals synthesized from the TiF_4 precursor is very much narrower, and is now limited by the machine broadening of 0.08° full width at half maximum. Importantly, the peak width of the mesoporous crystals, with 20 nm porosity, matches that of the non-templated anatase microcrystals, consistent with the porous particles being single-crystal domains; this correspondence is independent of pore size, consistent with a decoupling of crystallinity and surface area (Supplementary Information).

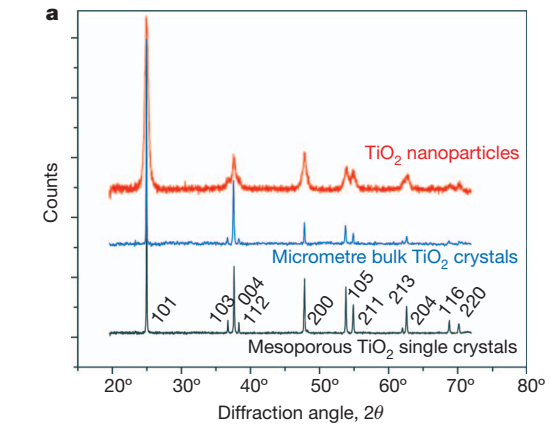
Electron diffraction taken from complete mesoporous particles produces the Laue diffraction patterns shown in Fig. 2c and e that are indexed to anatase TiO_2 with beam orientation consistent with the crystal symmetry of the Wulff construction. The distinct crystal shapes in Fig. 2b and d result from different reaction temperatures (described in detail in the Supplementary Information). Taken together with SEM characterization of particle size and symmetry, these results are entirely consistent with the mesoporous particles comprising a single underlying anatase crystal domain, with missing internal volume defined by the mesoporous template—we therefore refer to this material as a mesoporous single crystal (MSC), and it is, to the best of our knowledge, the first such example for anatase TiO_2.

We now return to the question of seeding density, which can be used not only to increase the relative number of crystals that grow in the template, but also to control their size, given that the seeded crystals grow in parallel, consuming all the available Ti precursor if the reaction is run to completion. In Fig. 3 we show how the number of TiO_2 seeds (N) and hence the average crystal volume (inversely proportional to N) can be varied over four orders of magnitude by varying the $TiCl_4$ seeding solution concentration from 15 μM to 15 mM. The Ti content in the seeded template, determined by inductively coupled plasma atomic emission spectroscopy (ICPOES), increases approximately linearly with $[TiCl_4]$ in this range (see ICPOES results and corresponding SEM images in the Supplementary Information). This result demonstrates explicitly how the MSC domain size is decoupled from the internal pore dimension imposed by the template morphology.

As an alternative to varying seed density, we can control both MSC size and shape via the TiF_4 hydrothermal reaction conditions. Reaction temperature, for example, influences both the aspect ratio and lateral size of the MSCs for a fixed template morphology, seed density and reaction time. We can adjust the internal mesopore size by replicating silica bead templates of different diameter d, such that the surface area (in square metres per gram) of the quasi-close-packed templates is inversely proportional to d. We show MSC pore morphologies replicating bead diameters between 20 nm and 250 nm in the Supplementary Information. Control over the mesopore dimension is important not only for increasing specific surface area, but also for optimizing diffusion of species through the pore structure in, for example, high-efficiency dye-sensitized solar cells (SSCs) containing viscous liquid electrolytes or solid-state hole-conductors and for long-range ion diffusion in batteries[2].

We have demonstrated a synthesis protocol for anatase TiO_2 MSCs with independently tunable crystal and pore size. We now investigate whether they genuinely exhibit superior electronic properties in

©2013 Macmillan Publishers Limited. All rights reserved



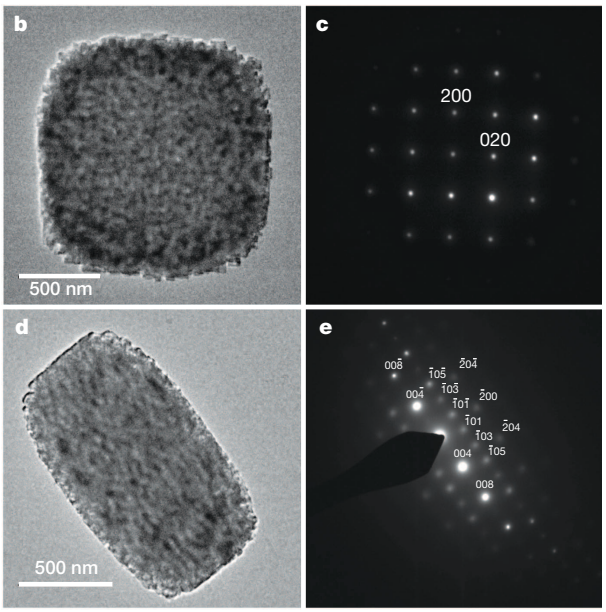


Figure 2 | X-ray and electron diffraction of mesoporous TiO_2 crystals. **a**, X-ray diffraction patterns of an ensemble of mesoporous TiO_2 crystals with 20-nm pore size and the equivalent non-templated microcrystals, compared to an ensemble of anatase nanoparticles (Dyesol DSL 18NR-T, 20-nm average particle size). **b** and **c**, Transmission electron micrograph (**b**) and electron diffraction Laue pattern (**c**) collected from a complete mesoporous crystal (synthesized at 190 °C, 40 mM TiF_4) assigned to anatase TiO_2 with [001] beam incidence. **d** and **e**, Equivalent image and Laue pattern for crystals synthesized at 130 °C, 120 mM TiF_4, indexed with [010] or [100] incidence.

comparison to mesoporous films constructed from dried or sintered nanoparticles. We first measure the electronic conductivity through an isolated MSC and compare it to that of TiO_2 nanoparticle films, summarizing the results in Fig. 4a. Before and after annealing at 500 °C, the MSC exhibits conductivity just under three and two orders of magnitude higher, respectively, across an approximately 500-nm electrode gap. The nanoparticle film conductivity increases far more upon sintering than does that of the MSC, rationalized by enhanced electronic contact through neck growth between the nanoparticles—a process now less relevant to MSCs because of pre-existing long-range crystal order[3,26]. Notably, the as-cast MSC is more conductive than the sintered nanoparticle film.

We also estimate the electronic mobility in MSC films, for which we have recently developed the requisite technique, which we call 'transient mobility spectroscopy'. Briefly, we measure the time-resolved conductivity and photoinduced charge absorption simultaneously to extract the charge density and hence mobility. In this configuration, the electrode gap is approximately 300 μm, so the measured MSC mobility is an effective mobility including the influence of interparticle boundaries.

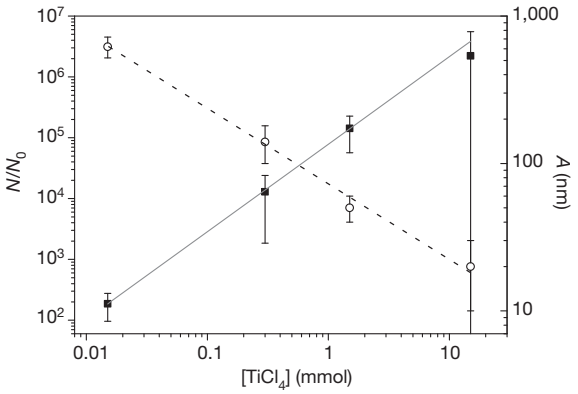


Figure 3 | MSC size control via nucleation seed density. MSC dimension A (circles) and inferred ratio of seed nuclei N to homogeneous nucleation N_0 as a function of $TiCl_4$ pre-treatment concentration (squares). Lines are drawn to guide the eye. Error bars for A represent one standard deviation of the measured sample distribution (minimum 100 samples). Primary hydrothermal reaction conditions were: 210 °C, 12 h, 20 mM TiF_4, in all cases.

Despite this, the MSC films exhibit mobility values over an order of magnitude higher than those of the nanoparticle films over a broad range of charge densities (Fig. 4b).

We have clearly demonstrated the advantageous electronic properties of the MSCs versus nanoparticle films, along with the unique property that MSC electron transport is highly effective even without any on-film heat treatment or sintering process. To capitalize on this commercially attractive property, we have used films of MSCs (300 nm crystal size to facilitate smooth film deposition) in low-temperature solid-state SSCs. All-solid-state SSCs, as opposed to those containing liquid electrolytes, suffer from fast electron–hole recombination, and the highest full sun power conversion efficiency reported for a solid-state SSC where all components are processed under 150 °C is a little over 1% (ref. 27). In Fig. 4c, we present current–voltage curves measured under simulated AM1.5 100 mW cm^{-2} sunlight for sub-150 °C processed SSCs using the organic dye sensitizer D102 or a mixed-halide organometallic perovskite sensitizer, $CH_3NH_3PbI_2Cl$ (refs 5, 28). The dye-sensitized cell achieves a 3% power conversion efficiency, comparable to the highest reported value of 4.1% for high-temperature-processed D102-sensitized anatase nanoparticles and vastly greater than the previous best low-temperature all-solid-state SSC[27]. MSCs coupled with the more effective perovskite light harvester deliver double the current and a power conversion efficiency of over 7%, higher than any reported low-temperature SSC architecture and competitive with high-temperature devices[5].

We have thus demonstrated a novel, highly effective, and versatile method for synthesizing micrometre-size mesoporous single-domain semiconductor particles based on crystal seeding in a sacrificial guiding template. Here we prove that this technique works very well for anatase TiO_2 in silica pores, but this method should be quite generally applicable to other functional ceramics and semiconductors. The complete decoupling of crystal size (atomic-scale order) from mesoscale porosity also provides an ideal system for fundamental investigation into the electronic processes in porous semiconductors. In addition, the accessibility of the exposed surface area opens up new possibilities for controlled surface treatments and solution-based core–shell synthesis, potentially decoupling long-range electronic and surface properties.

We have also demonstrated the promise held by MSCs for a specific application—SSCs. MSCs display higher conductivity and electron mobility than conventionally used nanocrystalline TiO_2, and the elimination of thermal sintering treatment enabled us to fabricate a solid-state SSC that attained a record 7.2% efficiency for sub-150 °C processing. This instantly opens up possibilities for multi-junction device fabrication, broadens substrate choice, and reduces manufacturing steps and costs for SSCs. More generally, incorporation of a

©2013 Macmillan Publishers Limited. All rights reserved

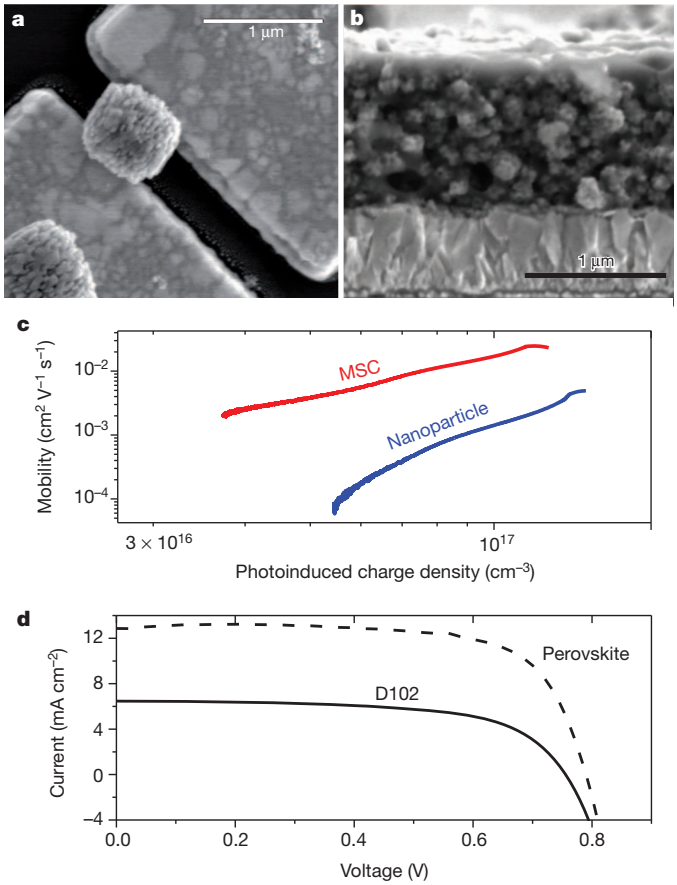


Figure 4 | Electronic properties and device performance of MSCs. a, SEM image of an MSC spanning a gold electrode gap. The conductivity of an isolated MSC is 4.9×10^{-6} S cm^{-1} before sintering at 500 °C and 1.5×10^{-5} S cm^{-1} afterwards. The conductivity of TiO_2 nanoparticle films before sintering at 500 °C is 8.4×10^{-9} S cm^{-1}and 2.2×10^{-7} S cm^{-1}afterwards. **b**, SEM cross-section of a solid-state dye-sensitized solar cell comprised of MSCs. **c**, Mobility dependence on photoinduced charge density for MSC and nanoparticle films measured via transient mobility spectroscopy. **d**, Current–voltage curves of all-low-temperature solid-state SSCs using MSC photoanodes with either the D102 dye or the perovskite sensitizer. The D102 device shows an efficiency of 3.11%, a short-circuit current of 6.47 mA cm^{-2}, an open-circuit voltage of 0.76 V and a fill factor of 0.63. The perovskite device shows an efficiency of 7.29%, a short-circuit current of 12.86 mA cm^{-2}, an open-circuit voltage of 0.79 V and a fill factor of 0.70.

variety of MSCs into active device layers should lead to great advances in operational efficiency of mesostructured solar cells, arising from rapid charge extraction in the inorganic phase, and improved battery operation due to enhanced charging and discharging rates.

METHODS SUMMARY

Mesoporous silica templates. Silica spheres were synthesized according to ref. 29 and collected by centrifugation. The template was seeded by immersion in aqueous $TiCl_4$ (15 μM to 15 mM) at 70 °C for 60 min followed by thorough rinsing with 1 litre H_2O.

TiO_2 MSC synthesis. Titanium tetrafluoride was dissolved (20 mM to 400 mM) in water (pH 2.1) to which was added the ionic liquid 1-methylimidazolium tetrafluoroborate (180 mM), and 650 mg of the pre-treated silica template for 50 ml of TiF_4 solution in a 125-ml-volume autoclave lined with Teflon (PTFE, poly(tetrafluroethane); Parr Instruments). The silica template was selectively etched in 2 M aqueous NaOH at 80 °C for 60 min in a polypropylene beaker and the TiO_2 product collected by centrifugation (3,000 r.p.m. for 60 min) and washed several times in water and ethanol.

Low-temperature solar cell assembly. Solid-state dye-sensitized solar cells were fabricated following the general process described in detail elsewhere[30]. A TiO_2 compact layer was first deposited by spin-coating a titanium isopropoxide solution in ethanol and dried at 150 °C. TiO_2 MSCs were deposited from suspension in ethanol containing 15 mol.% titanium diisopropoxide bis(acetylacetonate) relative to TiO_2, before sensitization with the D102 organic dye or coating with the perovskite precursor solution[5]. A Spiro-OMeTAD (2,2′7,7′-tetrakis-(N,N-dipmethoxyphenylamine)9,9′-spirobifluorene) hole transporter layer was deposited by spin-coating from cholorobenzene solutions containing tert-butylpyridine and lithium-bis(trifluoromethanesulphonyl)imide (TFSI) additives. 200-nm-thick silver counter electrodes were thermally evaporated over the hole transporter.

Characterization methods. Template and crystal-replica morphologies were investigated by SEM (Hitachi S-4300) at an accelerating voltage of 5 kV after platinum coating samples prepared by drop casting on glass substrates. Transmission electron microscope measurements were made with a Jeol2010 microscope operating at 200 kV. Dye desorption–adsorption measurements were conducted by sensitizing a thin layer of TiO_2 in the ruthenium dye N719 for 12 h. The dye was desorbed and its concentration evaluated from the peak optical absorption at 500 nm. Current–voltage characteristics of solar cells were measured under AM1.5 light at 100 mW cm^{-2} generated using an ABET Technologies solar simulator calibrated using a National Renewable Energy Laboratories (NREL) calibrated silicon reference cell with a KG5 filter to minimize spectral mismatch. Current–voltage curves were measured with a Keithley 2400 unit. All devices were masked with a metal aperture to define the active area and to eliminate edge effects. MSC conductivity measurements were taken using two-probe contact to gold electrodes with a single MSC crystal bridging the gap between them.

Full Methods and any associated references are available in the online version of the paper.

Received 11 December 2012; accepted 23 January 2013.
Published online 6 March 2013.

1. Weickert, J., Dunbar, R. B., Hesse, H. C., Wiedemann, W. & Schmidt-Mende, L. Nanostructured organic and hybrid solar cells. *Adv. Mater.* **23,** 1810–1828 (2011).
2. Yella, A. *et al.* Porphyrin-sensitized solar cells with cobalt (II/III) based redox electrolyte exceed 12 percent efficiency. *Science* **334,** 629–634 (2011).
3. O'Regan, B. & Grätzel, M. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2. *Nature* **353,** 737–740 (1991).
4. Grätzel, M. Photoelectrochemical cells. *Nature* **414,** 338–344 (2001).
5. Lee, M. M., Teuscher, J., Miyasaka, T., Murakami, T. N. & Snaith, H. J. Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites. *Science* **338,** 643–647 (2012).
6. Docampo, P., Guldin, S., Steiner, U. & Snaith, H. J. Charge transport limitations in self-assembled TiO_2 photoanodes for solid-state dye-sensitized solar cells. *J. Phys. Chem. Lett.* http://dx.doi.org/10.1021/jz400084n (in the press).
7. Schüth, F. Endo- and exotemplating to create high-surface-area inorganic materials. *Angew. Chem. Int. Edn* **42,** 3604–3622 (2003).
8. Lu, A.-H. & Schüth, F. Nanocasting: a versatile strategy for creating nanostructured porous materials. *Adv. Mater.* **18,** 1793–1805 (2006).
9. Dickinson, C. *et al.* Formation mechanism of porous single-crystal Cr_2O_3 and Co_3O_4 templated by mesoporous silica. *Chem. Mater.* **18,** 3088–3095 (2006).
10. Tian, B. *et al.* General synthesis of ordered crystallized metal oxide nanoarrays replicated by microwave-digested mesoporous silica. *Adv. Mater.* **15,** 1370–1374 (2003).
11. Yue, W. & Zhou, W. Synthesis of porous single crystals of metal oxides via a solid–liquid route. *Chem. Mater.* **19,** 2359–2363 (2007).
12. Yang, P., Zhao, D., Margolese, D. I., Chmelka, B. F. & Stucky, G. D. Generalised synthesis of large-pore mesoporous metal oxides with semicrystalline frameworks. *Nature* **396,** 152–155 (1998).
13. Li, D., Zhou, H. & Honma, I. Design and synthesis of self-ordered mesoporous nanocomposite through controlled in-situ crystallization. *Nature Mater.* **3,** 65–72 (2004).
14. Yu, J. C., Wang, X. & Fu, X. Pore-wall chemistry and photocatalytic activity of mesoporous titania molecular sieve films. *Chem. Mater.* **16,** 1523–1530 (2004).
15. Kondo, J. N. & Domen, K. Crystallization of mesoporous metal oxides. *Chem. Mater.* **20,** 835–847 (2008).
16. Yue, W. & Zhou, W. Crystalline mesoporous metal oxide. *Prog. Nat. Sci.* **18,** 1329–1338 (2008).
17. Jiao, K. *et al.* Growth of porous single-crystal Cr_2O_3 in a 3D mesopore system. *Chem. Commun.* 5618–5620 (2005).
18. Arora, H. *et al.* Block copolymer self-assembly-directed single-crystal homo- and heteroepitaxial nanostructures. *Science* **330,** 214–219 (2010).
19. Yue, W. *et al.* Mesoporous monocrystalline TiO_2 and its solid-state electrochemical properties. *Chem. Mater.* **21,** 2540–2546 (2009).
20. Bian, Z. *et al.* Single-crystal-like titania mesocages. *Angew. Chem. Int. Edn* **50,** 1105–1108 (2011).
21. Finnemore, A. S. *et al.* Nanostructured calcite single crystals with gyroid morphologies. *Adv. Mater.* **21,** 3928–3932 (2009).
22. Crossland, E. J. W. *et al.* A bicontinuous double gyroid dye-sensitized solar cell. *Nano Lett.* **9,** 2807–2812 (2009).

©2013 Macmillan Publishers Limited. All rights reserved

23. Yang, H. G. *et al.* Anatase TiO_2 single crystals with a large percentage of reactive facets. *Nature* **453,** 638–641 (2008).
24. Zhang, D., Li, G., Yang, X. & Yu, J. C. A micrometer-size TiO_2 single-crystal photocatalyst with remarkable 80% level of reactive facets. *Chem. Commun.* 4381–4383 (2009).
25. Liu, G., Yu, J. C., Lu, G. Q. M. & Cheng, H.-M. Crystal facet engineering of semiconductor photocatalysts: motivations, advances and unique properties. *Chem. Commun.* **47,** 6763–6783 (2011).
26. Nakade, S. *et al.* Dependence of TiO_2 nanoparticle preparation methods and annealing temperature on the efficiency of dye-sensitized solar cells. *J. Phys. Chem. B* **106,** 10004–10010 (2002).
27. Jiang, C. Y. *et al.* Low temperature processing solid-state dye sensitized solar cells. *Appl. Phys. Lett.* **100,** 113901 (2012).
28. Schmidt-Mende, L. *et al.* Organic dye for highly efficient solid-state dye-sensitized solar cells. *Adv. Mater.* **17,** 813–815 (2005).
29. Bogush, G. H., Tracy, M. A. & Zukoski, C. Z. IV Preparation of monodisperse silica particles: control of size and mass fraction. *J. Non-Cryst. Solids* **104,** 95–106 (1988).
30. Docampo, P. *et al.* Control of solid-state dye-sensitized solar cell performance by block-copolymer-directed TiO_2 synthesis. *Adv. Funct. Mater.* **20,** 1787–1796 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was funded by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 246124 of the SANS project, the European Research Council (HYPER project number 279881), the Rhodes Trust, the Engineering and Physical Sciences Research Council, and the Government of the Republic of Trinidad and Tobago. We thank C. Ducati for help with indexing of electron diffraction patterns.

Author Contributions E.J.W.C. and H.J.S. conceived the idea of the project. E.J.W.C. devised and performed materials synthesis and characterization. N.N. and V.S. fabricated and characterized solar cells and optoelectronic devices. T.L. and J.A.A.-W. contributed to electronic mobility measurements. E.J.W.C., H.J.S. and V.S. wrote the manuscript. All authors commented on the manuscript. H.J.S. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.J.S. (h.snaith1@physics.ox.ac.uk).

©2013 **Macmillan Publishers Limited. All rights reserved**

METHODS

Preparation of mesoporous silica templates. Silica spheres were synthesized according to ref. 29. For example, 50-nm spheres were synthesized by adding 32 ml H_2O (Millipore 18M Omega), 18 ml ammonium hydroxide (30%, Sigma Aldrich) and 99 ml tetraethyl orthosilicate (TEOS, Sigma Aldrich) to 750 ml ethanol and stirred at 700 r.p.m. for 24 h. A quasi-close-packed bead template was prepared by centrifugation of the reaction solution (7,000 r.p.m., 5 h) leaving a translucent solid. The (unwashed) solid was sintered at 500 °C for 30 min (ramping time 150 min).

Pre-seeding of silica templates. An aqueous $TiCl_4$ solution (15 μM to 15 mM) was prepared by aqueous dilution of an aqueous 2 M $TiCl_4$ stock solution (stock preparation: 100 ml of $TiCl_4$ (Sigma Aldrich) was added to 350 ml H_2O containing 1 ml of 35% HCl in an ice bath). 5 g of the sintered template was immersed in 33 ml of aqueous $TiCl_4$ and held at 70 °C for 60 min followed by thorough rinsing with 1 litre of H_2O. The dried template was resintered at 500 °C for 30 min (ramping time 150 min). The ratio of titanium to silicon (weight/weight) present in the treated template was determined by inductively coupled plasma optical emission spectrometry after microwave digestion in hydrofluoric acid (ICPOES, Exeter Analytical).

TiO_2 MSC synthesis. Mesoporous TiO_2 crystals were synthesized hydrothermally using a modified version of the procedure reported in ref. 24. Titanium tetrafluoride (TiF_4, Alfa Aesar) was dissolved in 18 MΩ H_2O (at concentrations from 20 mM to 400 mM) after first adjusting the pH to 2.1 by addition of 37% HCl (200 μl in 250 ml). 1.67 ml of the ionic liquid 1-methylimidazolium tetrafluoroborate, and 650 mg of the pre-treated silica template were added to 50 ml of TiF_4 solution in a 125-ml-volume Teflon-lined autoclave (Parr Instruments). The sealed vessel was brought to the target reaction temperature (100–210 °C) in a fan-assisted oven for the period of the reaction before cooling in air over approximately 1 h. The template product was collected as a solid white powder settled at the base of the vessel and rinsed with H_2O by vacuum filtration on 1.2-μm paper (Whatman). The silica template was selectively etched in aqueous 2 M NaOH at 80 °C for 60 min in a polypropylene beaker. The remaining TiO_2 was collected by centrifugation (3,000 r.p.m. for 60 min) and washed twice with 50 ml H_2O, and twice with 50 ml ethanol.

Low-temperature solar cell assembly. Solid-state dye-sensitized solar cells were fabricated following the general process described in detail elsewhere[30], and we describe only deviations from that process here. A TiO_2 compact layer was first deposited by spin-coating on a clean glass substrate coated with fluorine-doped tin oxide. The compact layer spin-coating solution was prepared by adding 0.71 g titanium isopropoxide and 0.07 g HCl to 8 ml of ethanol. After drying the compact layer at 150 °C for thirty minutes, the TiO_2 MSC mesostructure was deposited by spin-coating a solution of MSCs (300 nm crystal size, 50 nm pore size) dispersed in ethanol at 10% concentration by weight with the addition of 15 mol.% titanium diisopropoxide bis(acetylacetonate) relative to TiO_2. The substrates were again dried at 150 °C and subsequently immersed in the sensitizing solution containing D102 organic dye for one hour, or coated with the perovskite precursor solution described elsewhere[5]. After sensitization the Spiro-OMeTAD hole transporter layer was deposited by spin-coating from chlorobenzene at a concentration of 180 mg ml^{-1}, along with tert-butylpyridine and lithium-TFSI additives. 200-nm-thick silver counter electrodes were thermally evaporated over the hole transporter.

Characterization methods. Template and crystal-replica morphologies were investigated by SEM (Hitachi S-4300) at an accelerating voltage of 5 kV after platinum coating (3 nm) samples prepared by drop casting on glass substrates. Transmission electron microscope measurements were made with a Jeol2010 microscope operating at 200 kV. Dye desorption–adsorption measurements were conducted by sensitizing a thin layer of TiO_2 by immersion in a 10 mM solution (1:1 by volume of acetonitrile and tert-butanol) of the ruthenium dye N719 for 12 h. The dye was desorbed after thorough rinsing of the film in acetonitrile using a known volume of 200 mM aqueous KOH and the concentration evaluated from the peak absorption at 500 nm assuming a molar absorptivity coefficient of 14,000 M^{-1} cm^{-1} and effective area per molecule[31] of 1.8 nm^2. To obtain the current–voltage characteristics of low-temperature MSC-based solar cells, AM1.5 light at 100 mW cm^{-2} was generated using an ABET Sun 2000 class AAB solar simulator calibrated using an NREL calibrated silicon reference cell with a KG5 filter to minimize spectral mismatch. Current–voltage curves were measured with a Keithley 2400 unit. The solar cells were masked with a metal aperture to define the active area and to eliminate any edge effects. MSC conductivity measurements were taken using two-probe contact to gold electrodes with a single MSC bridging the gap between. To accomplish this, several pairs of gold electrodes were fabricated on silicon dioxide/silicon substrates using electron beam lithography, such that the electrodes could be contacted macroscopically, but narrowed to a width of 2.5 μm with a separating gap of 500 nm (see Fig. 4a). Subsequently, MSCs dissolved in ethanol were dropcast on the silicon substrate, dried and examined under an SEM to confirm the precise geometric parameters of any gap-bridging crystals present. Transient mobility spectroscopy will be described in detail elsewhere. Briefly, pump–probe spectroscopy is used to monitor the absorption signature of the photo-injected charge in the device. Conductivity devices are fabricated as described above with C106 sensitizing dye and no hole transporter. The pump wavelength is 550 nm to excite the dye molecules and induce photo-injection of electrons, whereas the probe wavelength of 900 nm measures the population of dye cations; simultaneously, the in-plane conductivity is measured. This allows the extraction of conductivity and photoinduced charge density as a function of time, from which the charge mobility can be calculated.

31. Zakeeruddin, S. M. *et al.* Design, synthesis, and application of amphiphilic ruthenium polypyridyl photosensitizers in solar cells based on nanocrystalline TiO_2 films. *Langmuir* **18,** 952–954 (2002).

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature11908

Water and hydrogen are immiscible in Earth's mantle

Enikő Bali[1,2], Andreas Audétat[1] & Hans Keppler[1]

In the deep, chemically reducing parts of Earth's mantle[1], hydrous fluids contain significant amounts of molecular hydrogen (H_2). Thermodynamic models of fluids in Earth's mantle so far have always assumed that molecular hydrogen and water are completely miscible. Here we show experimental evidence that water and hydrogen can coexist as two separate, immiscible phases. Immiscibility between water and hydrogen may be the cause of the formation of enigmatic, ultra-reducing domains in the mantle that contain moissanite (SiC) and other phases indicative of extremely reducing conditions[2,3]. Moreover, the immiscibility between water and hydrogen may provide a mechanism for the rapid oxidation of Earth's upper mantle immediately following core formation[4].

The oxygen fugacity of Earth's mantle controls the composition of volcanic gases and is therefore a key parameter in the evolution of Earth's atmosphere[5,6]. In general, it decreases from relatively oxidizing conditions (close to the fayalite–magnetite–quartz, FMQ, buffer) in the uppermost mantle to reducing conditions (defined by the iron–wüstite, Fe–FeO, buffer) at depths ≥200 km (ref. 1). This change in redox conditions is mostly due to the intrinsic pressure dependence of the equilibria between Fe^{2+} and Fe^{3+} in mantle minerals, and manifests itself in a trend towards lower oxygen fugacities with increasing depths, as recorded in mantle xenoliths[1]. However, oxygen fugacity does not only vary with depth, but also with local tectonic settings[7]. Relatively oxidized conditions are found in the mantle wedge above subduction zones[8], where oxidized material from Earth's surface is recycled into the mantle. In contrast, kimberlite eruptions in cratonic areas sometimes tap ultra-reducing mantle domains[2,3] that contain moissanite (SiC), metallic silicon and iron silicide ($FeSi_2$). The origin of these ultra-reducing domains is enigmatic, as they appear to record oxygen fugacities that are about six orders of magnitude below iron–wüstite[9,10], which is widely believed to be the lower limit of oxygen fugacities in the mantle.

Water-rich, aqueous fluids occur locally in the mantle, in particular at places where hydrous phases decompose—for example, in subduction zones. Under oxidizing conditions close to the FMQ buffer, such fluids contain a few mole % of molecular H_2. However, close to the iron–wüstite buffer, the molar fraction of H_2 in aqueous fluids approaches 50%. H_2 and H_2O have always been considered to be completely miscible in Earth's mantle. This is because the critical curve in the H_2–H_2O system, which gives the maximum temperature at which H_2 and H_2O may coexist as two separate phases, was found[11] to be located between 375 and 380 °C in the pressure range 0.1 to 0.25 GPa, with a dT/dP slope of <50 °C GPa^{-1}. In this study, we show that H_2 and H_2O may coexist as two separate phases in Earth's mantle and we suggest that this immiscibility may have been an essential factor in Earth's redox evolution.

In the present work, we performed high-pressure experiments in which aqueous fluids at oxygen fugacities of the iron–wüstite buffer were trapped as synthetic fluid inclusions[12] in silicate minerals such as olivine or quartz. Figure 1 shows a typical run-product from an experiment at 2.6 GPa and 1,000 °C (see Supplementary Information for more examples). The olivine crystal contains numerous fluid inclusions that formed in healed fractures during the experiment. Closer inspection shows that there are three different types of fluid inclusions present: (1) light-coloured fluid inclusions with weak optical contrast to the surrounding olivine; (2) dark-coloured fluid inclusions with strong optical contrast; and (3) fluid inclusions with a vapour bubble. Raman spectroscopy (Fig. 1) shows that the light inclusions contain nearly pure liquid water, whereas the dark inclusions are filled with high-density hydrogen gas. The ubiquitous coexistence of fluid inclusions of types (1) and (2) in the samples demonstrates that two immiscible phases, one rich in water, the other one nearly pure hydrogen, coexisted during run conditions. Some inclusions are intermediate in composition between pure water and pure hydrogen. These inclusions formed by the accidental trapping of two coexisting phases in one cavity in the growing crystal (Fig. 1).

Although immiscibility between H_2 and H_2O was observed at all pressures below a certain temperature, it disappeared at high temperature, and the corresponding samples show only one type of fluid inclusion (Fig. 2). These inclusions contain a bubble of high-density hydrogen in a liquid of nearly pure water, and therefore represent an intermediate composition in the H_2–H_2O system. From these data, it was possible to infer the approximate location of the critical curve in the H_2–H_2O system (Fig. 3). This curve gives minimum temperatures up to which immiscibility may occur, because this temperature will

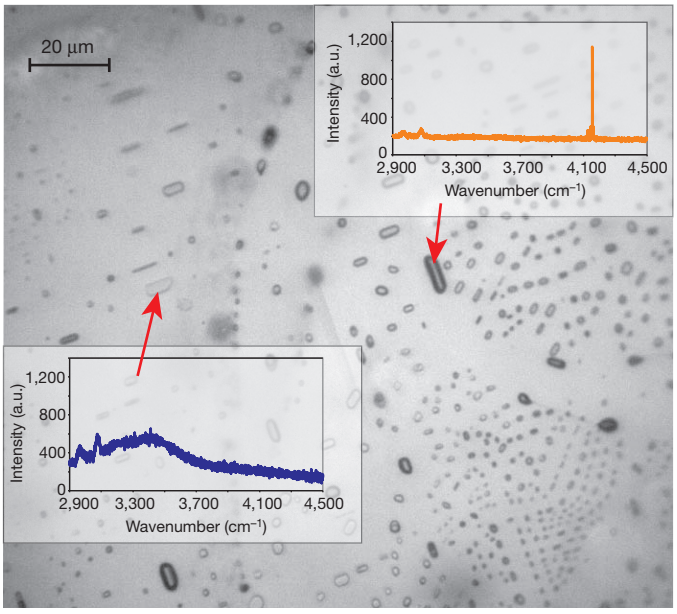


Figure 1 | Synthetic fluid inclusions in olivine formed at 1,000 °C and 2.6 GPa at Fe–FeO buffer conditions. Three different types of fluid inclusions are visible. The dark inclusions with strong optical contrast to the olivine contain high-density hydrogen according to the Raman spectra; an example is arrowed, and its Raman spectrum is shown inset (orange trace). The light-coloured inclusions contain liquid water; an example is arrowed, and its Raman spectrum is shown inset (blue trace). A few mixed inclusions, with a water-rich liquid phase and a hydrogen-rich bubble, are also seen. These observations suggest that under run conditions, a hydrogen-rich fluid phase coexisted with an immiscible water-rich fluid phase.

[1]Bayerisches Geoinstitut, Universität Bayreuth, 95440 Bayreuth, Germany. [2]Geowissenschaftliches Zentrum, Universität Göttingen, Goldschmidtstrasse 1, 37077 Göttingen, Germany.

©2013 Macmillan Publishers Limited. All rights reserved

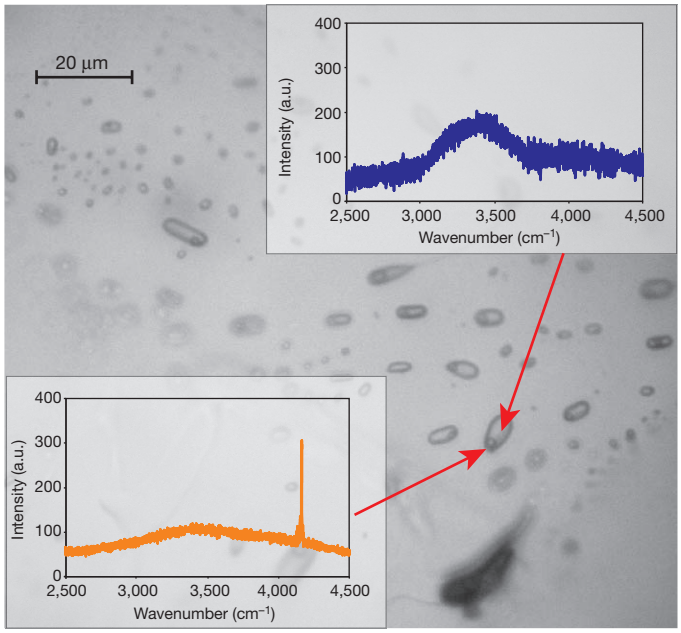


Figure 2 | Synthetic fluid inclusions in olivine formed at 950 °C and 2.3 GPa at Fe–FeO buffer conditions. Only one type of fluid inclusion is visible, consisting of a bubble of high-density hydrogen inside liquid water. Both phases are arrowed in an inclusion at lower right, and the respective Raman spectra are shown inset: orange trace, high-density hydrogen; and blue trace, liquid water. This observation suggests that at run conditions, water and hydrogen were fully miscible.

depend on the H_2/H_2O ratio and only one composition, buffered by iron–wüstite, was studied. Immiscibility may therefore occur even at somewhat (50–100 °C) higher temperatures than indicated in Fig. 3. However, because Fig. 3 is for iron–wüstite buffer conditions, which are probably representative for the deeper, reducing parts of the mantle, we will refer to it in the following discussions. The steep

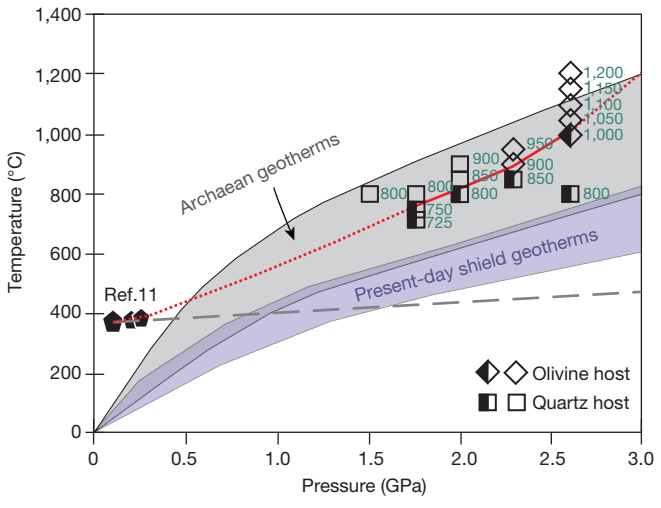


Figure 3 | Experimental data on the location of the critical curve in the H_2–H_2O system. Open symbols, only one type of fluid inclusion was observed in experiments. Black and white symbols, two types of fluid inclusion were observed, implying immiscibility between H_2 and H_2O. Numbers next to the symbols indicate the temperature of the experiment. Thick red line is the approximate location of the critical curve. Low-pressure data on the critical curve by Seward and Frank[11] are also shown (filled pentagons). Red dotted lines are the estimated location of the critical curve towards low and high pressures, based on the slope of the critical curve determined by our experiments. The modern shield geotherms are according to refs 14 and 15, the Archaean shield geotherms are from ref. 30.

increase of critical temperature with pressure is probably related to a positive excess volume of mixing on interaction of the polar H_2O molecules with the non-polar H_2. The experiments reported here were all carried out in the presence of quartz or olivine, so that the fluids will contain some dissolved silica and minor MgO—as in nature, where fluids in the upper mantle will coexist with olivine. However, we do not expect a significant shift of the critical curve relative to the pure H_2O–H_2 system, as the silica content of our supercritical fluids should have already been very low due to the strong decrease in silica solubility on dilution of water by non-polar molecules[13].

The composition of the coexisting fluid phases in the H_2O–H_2 system can also be roughly estimated from the fluid inclusion data. By measuring the size of the gas bubble in inclusions, one can estimate the bulk density of the fluid; the ratio of this number to the density of pure water at the given pressure and temperature conditions then gives a first approximation of the molar fraction of water in the fluid. These data suggest that at 2.3 GPa and 950 °C (above the critical curve) the fluid contains between 50 and 55 mol% H_2, whereas at 850 °C, the two coexisting phases appear to be water with less than 5 mol% of H_2 and nearly pure H_2 with less than 5 mol% water. In particular, the gas-rich inclusions show no clear evidence of the presence of water; however, there may be a thin film of water absorbed on the inner surface of the inclusion. Observations at other pressures also suggest very limited miscibility between H_2 and H_2O at 100 °C below the critical curve.

In Fig. 3, the inferred location of the critical curve is compared to a range of modern shield geotherms[14,15]. It is clear that immiscibility between hydrogen and water will occur in reducing, hydrous fluids in Earth's mantle. Such fluids may be produced by dehydration reactions; for example, the dehydration of hydrous minerals transported by deep subduction[16,17] into the reducing portions of the mantle, where water may be temporarily stored in hydrous phase A (a high-pressure Mg-silicate containing water). If these fluids then equilibrate with the oxygen fugacity of the surrounding bulk mantle, an immiscible hydrogen-rich gas phase will form. According to the data from fluid inclusions in our experiments (Fig. 1), this fluid phase will contain very little water. Owing to its very low density and low viscosity, it will rapidly migrate upwards. As the fluid migrates, the remaining water will be dissolved in the surrounding nominally anhydrous minerals that are normally water-undersaturated throughout most of the mantle. This water removal process is much more efficient than the oxidation of hydrogen by the traces of Fe^{3+} in the upper mantle (for a quantitative discussion, see Supplementary Information). On removal of residual water from the fluid, oxygen fugacity will decrease by orders of magnitude, such that moissanite, iron silicide and even metallic silicon become stable. Moissanite may form by the reaction $2MgSiO_3$ (enstatite) + C (diamond) = Mg_2SiO_4 (forsterite) + SiC (moissanite) + O_2 throughout a wide range of upper-mantle pressure and temperature conditions, if oxygen fugacity drops by about six log units below the iron–wüstite value[9,10]. We therefore suggest that the ultra-reducing mantle domains sampled by kimberlites are the results of deep dehydration reactions—for example, in old subducted material[16,17]—that produce an immiscible hydrogen-rich fluid phase.

The evolution of mantle oxidation state with time has been the subject of intense research, as it controls the composition of volcanic gases[5,6] as well as the behaviour of many trace elements during mantle melting, which may ultimately lead to economic ore deposits[18]. Mantle oxidation state also exerts a fundamental control on Earth's deep carbon cycle. The early Earth's mantle may have been poor in carbon[19], with most carbon residing in the atmosphere or in the core, because the solubility of reduced carbon (graphite or methane) in a magma ocean is very low[20]. Fluids present in the early mantle are thus expected to have been H_2O-dominated, in contrast to the CO_2-dominated fluids now present in the upper, oxidized mantle[21]. Numerous studies have attempted to detect a change in mantle oxidation state over geologic time, but failed to find any evidence for such an event in the past four billion years (refs 22–24). Rather, there is evidence that the oxidation

©2013 Macmillan Publishers Limited. All rights reserved

of Earth's upper mantle must have occurred very rapidly after core formation[4]. The mechanism that could lead to such a rapid oxidation, however, remains uncertain[25]. We suggest that this event was caused by degassing of nearly pure hydrogen that formed by immiscibility in H_2O–H_2 fluids coexisting with a reduced early mantle at an oxygen fugacity close to iron–wüstite, as expected after core formation.

Figure 3 shows that H_2–H_2O immiscibility is expected to occur even at the elevated geothermal gradients in the Archaean. If a hydrogen-rich fluid phase is extracted from the mantle, the residual reservoir must become oxidized. This process can be visualized by the reaction $2FeO + H_2O = Fe_2O_3 + H_2$, where FeO and Fe_2O_3 are the components of ferrous and ferric iron in the silicate minerals of the mantle, and where H_2 is rapidly removed from the system. The early, undegassed mantle probably contained much more water than the present mantle[26,27], and large amounts of water may have been released as free fluid on convective upwelling of a hydrous transition zone. Buoyant plumes of hydrogen gas produced by immiscibility in the H_2–H_2O system may have reached the surface. This process may have already happened at the time when core formation started. An attractive aspect of this model is that it oxidizes the mantle without much affecting its bulk or trace-element composition, as the solubility of nearly all elements (except perhaps some volatiles) in pure hydrogen is expected to be negligibly low. This model also provides a mechanism to flush the atmosphere for a limited period of time with gases that are much more reduced than what was previously believed to be possible, so that a reducing early atmosphere could have been stabilized for some time[28].

METHODS SUMMARY

Samples of fluids were trapped as synthetic fluid inclusions in quartz or olivine in piston-cylinder experiments. Small cubes of quartz or olivine were loaded with water into a platinum-rhodium capsule, which was sealed by arc welding[29] and enclosed in a larger capsule made of pure iron, which contained some FeO and water to buffer oxygen fugacity at the iron–wüstite buffer. The iron capsule was sealed mechanically on compression of the MgO-NaCl assembly. Experiments were carried out in an end-loaded piston cylinder apparatus for 24 h and quenched to room temperature in approximately 4 min while approximately 30% of the pressure was also released simultaneously. This method prevented the decrepitation (explosion) of most fluid inclusions and at the same time quartz or olivine recrystallization during the quench process was minimal. Sample capsules were cut open, the chips of olivine or quartz were removed and doubly polished. Fluid inclusions in the minerals were studied by optical microscopy and by confocal Raman spectroscopy (Horiba Jobin Yvon LabRAM HR800, 514 nm Ar laser with 200 mW output power, 1,800 mm^{-1} grating, 500 μm confocal pinhole, 50× long-distance objective).

Full Methods and any associated references are available in the online version of the paper.

Received 25 September 2012; accepted 15 January 2013.

1. Frost, D. J. & McCammon, C. A. The redox state of Earth's mantle. *Annu. Rev. Earth Planet. Sci.* **36,** 389–420 (2008).
2. Shiryaev, A. A., Griffin, W. L. & Stoyanov, E. Moissanite (SiC) from kimberlites: polytypes, trace elements, inclusions and speculations on origin. *Lithos* **122,** 152–164 (2011).
3. Trumbull, R. B. *et al.* The carbon isotope composition of natural SiC (moissanite) from the Earth's mantle: new discoveries from ophiolites. *Lithos* **113,** 612–620 (2009).
4. Trail, D., Watson, E. B. & Tailby, N. D. The oxidation state of Hadean magmas and implications for early Earth's atmosphere. *Nature* **480,** 79–82 (2011).
5. Holland, H. D. Volcanic gases, black smokers, and the Great Oxidation Event. *Geochim. Cosmochim. Acta* **66,** 3811–3826 (2002).
6. Gaillard, F., Scaillet, B. & Arndt, N. T. Atmospheric oxygenation caused by a change in volcanic degassing pressure. *Nature* **478,** 229–232 (2011).
7. Wood, B. J., Bryndzia, L. T. & Johnson, K. E. Mantle oxidation-state and its relationship to tectonic environment and fluid speciation. *Science* **248,** 337–345 (1990).
8. Kelley, K. A. & Cottrell, E. Water and the oxidation state of subduction zone magmas. *Science* **325,** 605–607 (2009).
9. Ulmer, G. C. *et al.* The redox stability of moissanite (SiC) compared with metal – metal oxide buffers at 1773 K and at pressures up to 90 kbar. *Neues Jb. Mineral. Abh.* **172,** 297–307 (1998).
10. Mathez, E. A., Fogel, R. A., Hutcheon, I. D. & Marshintsev, V. K. Carbon isotopic composition and origin of SiC from kimberlites of Yakutia, Russia. *Geochim. Cosmochim. Acta* **59,** 781–791 (1995).
11. Seward, T. M. & Franck, E. U. The system hydrogen-water up to 440 °C and 2500 bar pressure. *Ber. Bunsenges. Phys. Chem* **85,** 2–7 (1981).
12. Sterner, S. M. & Bodnar, R. J. Synthetic fluid inclusions in natural quartz. 1. Compositional types synthesized and applications to experimental geochemistry. *Geochim. Cosmochim. Acta* **48,** 2659–2668 (1984).
13. Walther, J. V. & Orville, P. M. The extraction-quench technique for determination of the thermodynamic properties of solute complexes: application to quartz solubility in fluid mixtures. *Am. Mineral.* **68,** 731–741 (1983).
14. Jaupart, C. & Mareschal, J. C. The thermal structure and thickness of continental roots. *Lithos* **48,** 93–114 (1999).
15. Rudnick, R. L. & Nyblade, A. A. in *Mantle Petrology: Field Observations and High-Pressure Experimentation: A Tribute to Francis R. (Joe) Boyd* (eds Fei, Y.-W., Bertka, C. M. & Mysen, B. O.) 3–12 (Spec. Publ. 6, Geochemical Soc., 1999).
16. Kawamoto, T. Hydrous phases and water transport in the subducting slab. *Rev. Mineral. Geochem.* **62,** 273–289 (2006).
17. Rupke, L. H., Morgan, J. P., Hort, M. & Connolly, J. A. D. Serpentine and the subduction zone water cycle. *Earth Planet. Sci. Lett.* **223,** 17–34 (2004).
18. Sillitoe, R. H. Characteristics and controls of the largest porphyry copper-gold and epithermal gold deposits in the circum-Pacific region. *Aust. J. Earth Sci.* **44,** 373–388 (1997).
19. Dasgupta, R. & Hirschmann, M. M. The deep carbon cycle and melting in Earth's interior. *Earth Planet. Sci. Lett.* **298,** 1–13 (2010).
20. Morizet, Y., Paris, M., Gaillard, F. & Scaillet, B. C-O-H fluid solubility in haplobasalt under reducing conditions: an experimental study. *Chem. Geol.* **279,** 1–16 (2010).
21. Andersen, T. & Neumann, E. R. Fluid inclusions in mantle xenoliths. *Lithos* **55,** 301–320 (2001).
22. Canil, D. Vanadium partitioning and the oxidation state of Archaean komatiite magmas. *Nature* **389,** 842–845 (1997).
23. Berry, A. J., Danyushevsky, L. V., O'Neill, H. S. C., Newville, M. & Sutton, S. R. Oxidation state of iron in komatiitic melt inclusions indicates hot Archaean mantle. *Nature* **455,** 960–963 (2008).
24. Li, Z. X. A. & Lee, C. T. A. The constancy of upper mantle fO_2 through time inferred from V/Sc ratios in basalts. *Earth Planet. Sci. Lett.* **228,** 483–493 (2004).
25. Williams, H. M., Wood, B. J., Wade, J., Frost, D. J. & Tuff, J. Isotopic evidence for internal oxidation of the Earth's mantle during accretion. *Earth Planet. Sci. Lett.* **321,** 54–63 (2012).
26. Marty, B. The origins and concentrations of water, carbon, nitrogen and noble gases on Earth. *Earth Planet. Sci. Lett.* **313,** 56–66 (2012).
27. Morbidelli, A. *et al.* Source regions and timescales for the delivery of water to the Earth. *Meteorit. Planet. Sci.* **35,** 1309–1320 (2002).
28. Tian, F., Toon, O. B., Pavlov, A. A. & De Sterck, H. A hydrogen-rich early Earth atmosphere. *Science* **308,** 1014–1017 (2005).
29. Audetat, A. & Bali, E. A new technique to seal volatile-rich samples into platinum capsules. *Eur. J. Mineral.* **22,** 23–27 (2010).
30. Mareschal, J. C. & Jaupart, C. in *Archean Geodynamics and Environments* (eds Benn, K., Mareschal, J. & Condie, K. C.) 61–73 (Geophys. Monograph Ser. 164, Am. Geophys. Union, 2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Schulze for help in sample preparation. This Letter has been improved by comments from D. Walker.

Author Contributions E.B. carried out all experiments reported in this Letter. A.A. assisted in the design and interpretation of the experiments. H.K. helped in measuring the Raman spectra and in the geological application of the data. The authors wrote the manuscript together.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.B. (enikoe.bali@geo.uni-goettingen.de)

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Starting materials. Fluid inclusions were trapped in inclusion-free quartz from Brazil at temperatures below 850 °C and in inclusion-free olivine from Dak Lac Province in Vietnam at higher temperatures. One or two cubes of quartz or fractured olivine crystals were loaded into $Pt_{95}Rh_5$ capsules together with 20–30 μl of distilled water. Capsules were 8 mm long having an outer diameter of 3.5 mm and an inner diameter of 3.2 mm. Capsules were prepared according to ref. 29. Before every experiment, capsules were held in a drying oven for at least a few hours (preferably overnight) to check for possible water leaks.

In order to buffer the oxygen fugacity at conditions equilibrium with Fe–FeO (iron–wüstite), iron capsules were machined from iron rods which had a purity of >99.5%. The length of the iron capsules was 10 mm, the outer diameter was 5 mm, and the inner diameter was 4.4 mm. The PtRh capsules were placed into these iron capsules with FeO and distilled water. These outer iron capsules were closed by pressurization during the experiment.

High-pressure experiments. The experiments were carried out in an end-loaded piston cylinder apparatus at the Bayerisches Geoinstitut, Bayreuth, and at the Department of Experimental and Applied Mineralogy, University of Göttingen. In both cases we used ½ inch MgO-NaCl assemblies, with stepped graphite heaters. The temperature was measured by S-type (Pt/Pt-Rh) thermocouple and monitored by a Eurotherm controller. A friction correction of 5% was applied.

Pressure and temperature were raised simultaneously along a fluid isochoric path. This method and the use of MgO-NaCl pressure cells led to minimal deformation of the capsules. All experiments were run for 24 h. Experiments were quenched in approximately 4–5 min. During this period, approximately 30% of the pressure was also released. After room temperature was reached the remaining pressure was lowered to room pressure in approximately 15–20 min. This ensured that a high percentage of the fluid inclusions survived quenching and decompression; moreover, our previous experience showed that during the 4 min quenching period, low-pressure and low-temperature recrystallization is minimal.

Investigation of run products. After the experiments, quartz and olivine crystals were retrieved from the capsules, mounted in epoxy and doubly polished. Fluid inclusions in the minerals were studied by optical microscopy and by confocal Raman spectroscopy. The optical observation of fluid inclusions was carried out by a Leica DMEP polarization microscope equipped with 4× to 60× objectives. Photomicrographs were taken by a Zeiss Axiocam ERC5 s camera coupled with Axiovision 4.8.2 imaging software. Raman spectroscopy was carried out with a Horiba Jobin Yvon LabRAM HR 800 Raman microscope with 800 mm focal length, 514 nm Ar laser with 200 mW output power, 1,800 mm^{-1} grating, 500 μm confocal pinhole, 50× long-distance objective with NA = 0.35, Peltier-cooled CCD detector with 1,024 × 256 pixels of 26 × 26 μm. Spectra were accumulated for 900 s. Spectral resolution is 3.5 cm^{-1}.

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature11906

CALHM1 ion channel mediates purinergic neurotransmission of sweet, bitter and umami tastes

Akiyuki Taruno[1]*, Valérie Vingtdeux[2]*, Makoto Ohmoto[3], Zhongming Ma[1], Gennady Dvoryanchikov[4], Ang Li[1]†, Leslie Adrien[2], Haitian Zhao[2], Sze Leung[5], Maria Abernethy[5], Jeremy Koppel[2], Peter Davies[2,6], Mortimer M. Civan[1,7], Nirupa Chaudhari[4,8], Ichiro Matsumoto[3], Göran Hellekant[5], Michael G. Tordoff[3], Philippe Marambaud[2] & J. Kevin Foskett[1,9]

Recognition of sweet, bitter and umami tastes requires the non-vesicular release from taste bud cells of ATP, which acts as a neurotransmitter to activate afferent neural gustatory pathways[1]. However, how ATP is released to fulfil this function is not fully understood. Here we show that calcium homeostasis modulator 1 (CALHM1), a voltage-gated ion channel[2,3], is indispensable for taste-stimuli-evoked ATP release from sweet-, bitter- and umami-sensing taste bud cells. *Calhm1* knockout mice have severely impaired perceptions of sweet, bitter and umami compounds, whereas their recognition of sour and salty tastes remains mostly normal. *Calhm1* deficiency affects taste perception without interfering with taste cell development or integrity. CALHM1 is expressed specifically in sweet/bitter/umami-sensing type II taste bud cells. Its heterologous expression induces a novel ATP permeability that releases ATP from cells in response to manipulations that activate the CALHM1 ion channel. Knockout of *Calhm1* strongly reduces voltage-gated currents in type II cells and taste-evoked ATP release from taste buds without affecting the excitability of taste cells by taste stimuli. Thus, CALHM1 is a voltage-gated ATP-release channel required for sweet, bitter and umami taste perception.

Tastes are sensed by dedicated receptor cells in taste buds, which are composed of three anatomically distinct types of cells: types I, II and III. Only type III cells, which sense sour tastes, have neuron-like features, including expression of neurotransmitter biosynthesis enzymes and synaptic vesicles at classical synapses with sensory nerve fibres[1]. Type II cells, which sense sweet, bitter and umami tastes, share a common signal-transduction pathway. However, they lack classical synaptic structures, yet transmit taste information to gustatory neurons by releasing ATP as a neurotransmitter[4,5]. Our work, described below, implicates CALHM1 as a critical component responsible for this ATP release.

Calhm1 (ref. 6) encodes the pore-forming subunit of a voltage-gated, non-selective, plasma membrane ion channel involved in neuronal excitability[2] and, potentially, the pathogenesis of Alzheimer's disease[6–8]. *Calhm1* expression was identified in primate taste buds[9], suggesting that CALHM1 may have physiological functions outside the brain. We confirmed that *Calhm1* was expressed in mouse taste buds but not in surrounding epithelium (Fig. 1a, b, e–g). To examine CALHM1 function, we generated a constitutive *Calhm1*$^{-/-}$ mouse and verified loss of *Calhm1* expression in taste buds (Fig. 1a, c). *Calhm1*$^{-/-}$ mice were viable and fertile, with no overt morphological abnormalities in their taste buds nor any altered expression of taste-related marker genes (Fig. 1a and Supplementary Fig. 1). Loss of *Calhm1* signal in taste buds of *Skn-1a* (also known as *Pou2f3*) knockout mice, from which type II cells are developmentally absent[10] (Fig. 1d and Supplementary Fig. 2), and the co-expression of *Calhm1* and *Trpm5* (Fig. 1h), demonstrated that *Calhm1* expression is confined to type II cells. Expression profiling by reverse-transcription PCR of pools of isolated type II and type III cells and individual taste cells also supported the confined expression of *Calhm1* to type II cells (Supplementary Fig. 3). *Calhm1* expression was observed not only in *Tas1r3*-expressing sweet and umami taste cells, but also in other type II cells, indicating that *Calhm1* is expressed in sweet, bitter and umami taste cells (Fig. 1i).

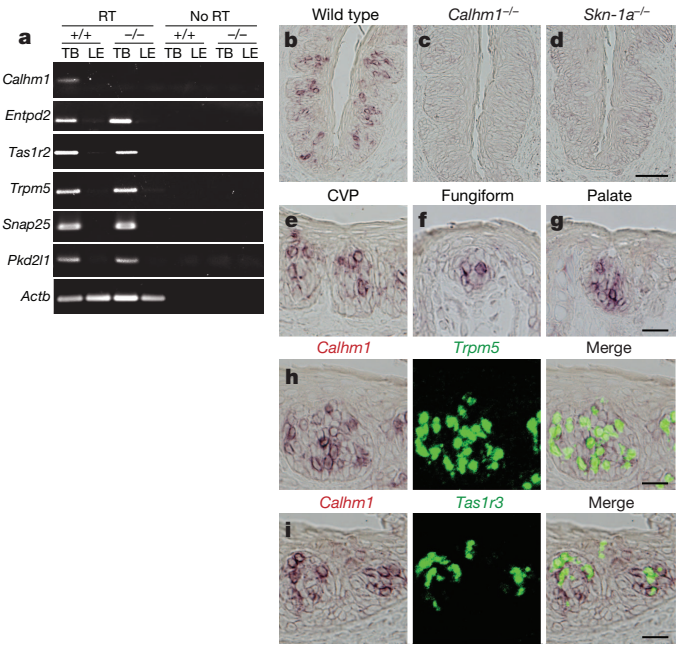


Figure 1 | CALHM1 is selectively expressed in type II taste bud cells. **a**, Reverse-transcription PCR of messenger RNA of *Calhm1*, *Actb* (β-actin) and taste cell marker genes from laser-microdissected circumvallate papillae (CVP) taste buds (TB) and lingual epithelium (LE) in wild-type (+/+) and *Calhm1* knockout (−/−) mouse tongues. RT, reverse transcriptase. **b**–**d**, *In situ* hybridization of *Calhm1* in CVP taste buds of wild-type (**b**), *Calhm1*$^{-/-}$ (**c**) and *Skn-1a*$^{-/-}$ (**d**) mice. Scale bar, 50 μm. **e**–**g**, *Calhm1* is expressed in subsets of CVP (**e**), fungiform (**f**) and palate (**g**) taste bud cells. **h**, Double-label *in situ* hybridization directly illustrates cellular co-expression of *Calhm1* and *Trpm5* in CVP taste buds. Most cells expressing *Trpm5* also express *Calhm1*, with *Calhm1* expression absent from *Trpm5*-negative cells. **i**, CVP taste buds illustrating that *Tas1r3* is expressed in a subset of *Calhm1*-positive cells. Scale bars for **e**–**i**, 20 μm.

[1]Department of Physiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [2]Litwin-Zucker Research Center for the Study of Alzheimer's Disease, The Feinstein Institute for Medical Research, Manhasset, New York 11030, USA. [3]Monell Chemical Senses Center, Philadelphia, Pennsylvania 19104, USA. [4]Department of Physiology and Biophysics, Miller School of Medicine, University of Miami, Miami, Florida 33136, USA. [5]Department of Biomedical Sciences, Medical School, University of Minnesota Duluth, Duluth, Minnesota 55812, USA. [6]Department of Pathology, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [7]Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [8]Program in Neurosciences, Miller School of Medicine, University of Miami, Miami, Florida 33136, USA. [9]Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. †Present address: Department of Anatomy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong.
*These authors contributed equally to this work.

©2013 Macmillan Publishers Limited. All rights reserved

We used two behavioural tests and nerve recordings to evaluate the effect of the loss of CALHM1 expression on taste perception. Two-bottle preference tests revealed that *Calhm1*$^{-/-}$ mice had a nearly complete loss both of preference for sweet and umami compounds and of avoidance of bitter compounds (Fig. 2a, Supplementary Fig. 4a, c, d and Supplementary Tables 1 and 2). In contrast, there were little or no differences between wild-type and *Calhm1*$^{-/-}$ mice in their responses to salty and sour compounds (Fig. 2a, Supplementary Fig. 4a and Supplementary Tables 1 and 2). Brief-access taste tests reproduced these phenotypes (Fig. 2b, Supplementary Fig. 4b and Supplementary Table 3), demonstrating that the taste phenotype in *Calhm1*$^{-/-}$ mice is a sensory defect. Recordings of the whole chorda tympani nerve revealed strongly reduced responses to sweet, bitter and umami compounds in *Calhm1*$^{-/-}$ mice, whereas responses to NaCl and the acid compounds were not different from wild type (Fig. 2c and Supplementary Fig. 5), indicating that the sensory defect in *Calhm1*$^{-/-}$ mice was due to the knockout of CALHM1 function in the peripheral taste cell system. Together, these results indicate that CALHM1 has a crucial role in taste-sensing type II cells[1,11] (see Supplementary Information for further discussion).

Type II cells signal to the nervous system by non-vesicular ATP release to nearby afferent gustatory nerves[5,12–14]. Although connexin hemichannels and pannexin 1 have been proposed, the molecular identity of the ATP-release mechanism remains uncertain[1,12,13,15]. The CALHM1 pore diameter is ~14 Å (ref. 3), similar to that estimated for connexins. We therefore examined the possibility that CALHM1

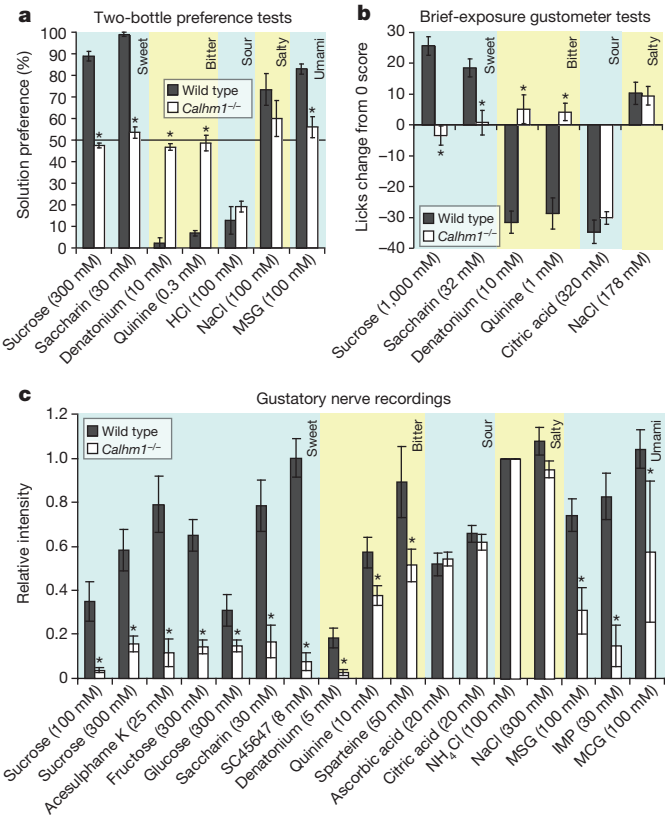


Figure 2 | CALHM1 is essential for sweet, bitter and umami taste perception. **a**, **b**, Mean preference percentage (taste compound versus water) from 48-h, two-bottle preference tests (**a**) and brief-access lick scores (**b**) for indicated compounds in *Calhm1*$^{-/-}$ mice and wild-type littermates. Error bars, s.e. (8–12 mice per group, 4–6 months old); *$P < 0.01$ (*post hoc* Bonferroni's test (**a**) and Student's *t*-test (**b**)). **c**, Summary of responses from whole-chorda-tympani nerve recordings stimulated with indicated compounds and normalized to response to NH_4Cl from wild-type ($n = 7$) and *Calhm1*$^{-/-}$ ($n = 8$) mice. MSG, monosodium glutamate; IMP, inosine 5′-monophosphate; MCG, monocalcium di-L-glutamate. Error bars, s.e.; *$P < 0.05$ (Student's *t*-test).

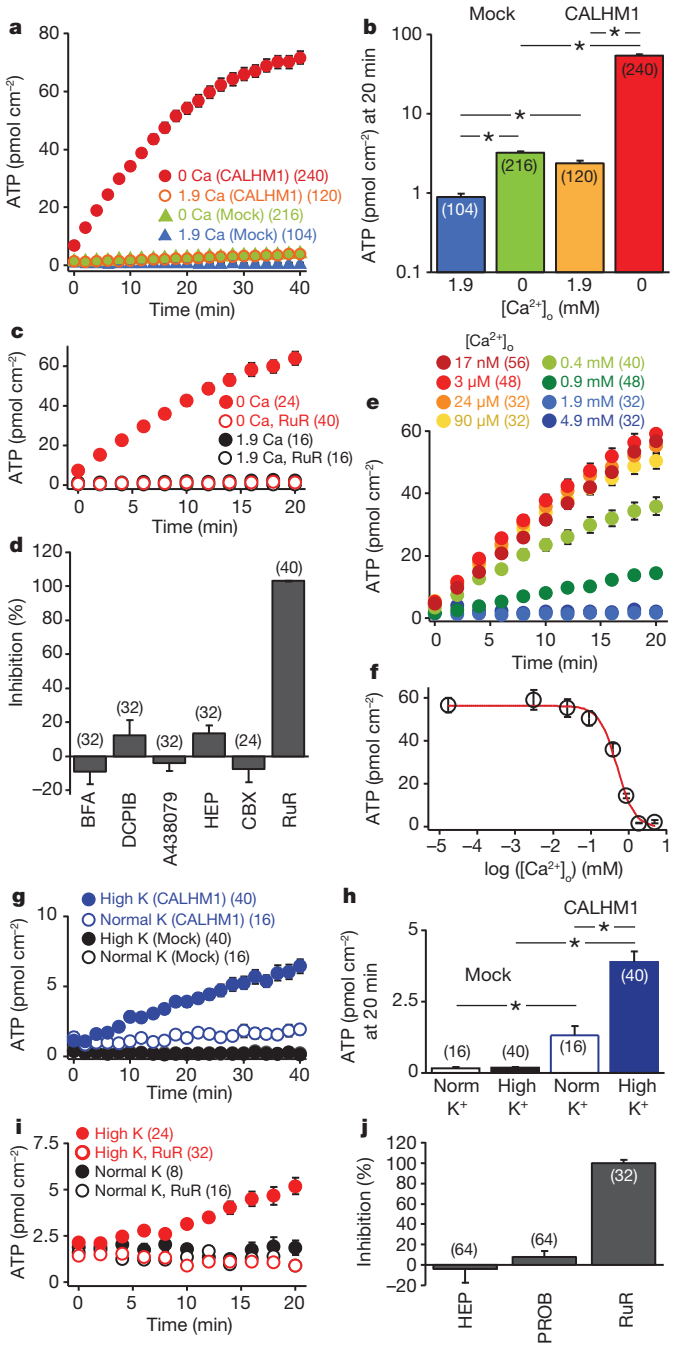


Figure 3 | CALHM1 mediates ATP release. **a**, Time courses of extracellular ATP levels due to release from mock- and hCALHM1-transfected HeLa cells exposed to normal (1.9 mM) or essentially zero (17 nM) $[Ca^{2+}]_o$. **b**, Summary of extracellular ATP levels at 20 min in **a**. **c**, Low-$[Ca^{2+}]_o$-induced ATP release in CALHM1-expressing cells is abolished by ruthenium red (RuR). **d**, Effects on low-$[Ca^{2+}]_o$-stimulated ATP release from hCALHM1 cells of 0.5 µg ml^{-1} brefeldin A (BFA), 10 µM DCPIB, 3 µM A438079, 1 mM 1-heptanol (HEP), 30 µM carbenoxolone (CBX) and 20 µM RuR. **e**, Time courses of ATP release from hCALHM1 cells induced by various $[Ca^{2+}]_o$. **f**, ATP levels at 20 min in **e** plotted against $[Ca^{2+}]_o$ and fitted to a Hill equation. **g**, Depolarization by high $[K^+]_o$ (117.5 mM K^+) induces ATP release specifically from hCALHM1 cells. **h**, ATP levels at 20 min in **g**. **i**, Depolarization-induced ATP release from CALHM1-expressing cells is abolished by RuR. **j**, Pharmacological sensitivities of depolarization-induced ATP release from hCALHM1 cells: 1 mM HEP, 1 mM probenecid (PROB), 20 µM RuR. Number of wells shown in parentheses throughout. Error bars, s.e. (where visible); *$P < 0.01$ (Student's *t*-test).

©2013 Macmillan Publishers Limited. All rights reserved

mediates ATP release. The CALHM1 ion channel can be activated by membrane depolarization or reduction of extracellular Ca^{2+} concentration[2] ($[Ca^{2+}]_o$). We first exploited the latter mechanism to activate CALHM1 in heterologous expression systems to determine whether CALHM1 can form an ATP-release channel. Decreasing $[Ca^{2+}]_o$ from 1.9 mM to essentially zero (17 nM) activated ATP release from human CALHM1 (hCALHM1)-expressing HeLa cells, whereas little ATP efflux was induced in mock-transfected cells (Fig. 3a, b). Similar low-$[Ca^{2+}]_o$-induced ATP release was observed in hCALHM1-expressing COS-1 cells and *Xenopus* oocytes (Supplementary Fig. 6). Neither CALHM1 expression nor low-$[Ca^{2+}]_o$ exposure caused cell damage (Supplementary Fig. 7). Involvement of other possible mechanisms[16] was ruled out because ATP release was unaffected by brefeldin A (vesicular release), DCPIB (volume-sensitive Cl^- channels), A438079 (P2X7 receptors), heptanol (connexin hemichannels) or carbenoxolone (pannexins and connexins at 30 μM) (Fig. 3d). In contrast, ruthenium red (RuR), which inhibits CALHM1 ion currents[2], abolished low-$[Ca^{2+}]_o$-evoked ATP release from hCALHM1-expressing cells (Fig. 3c, d). Thus, CALHM1 expression induces a novel ATP permeability.

CALHM1 ion channel gating is regulated by $[Ca^{2+}]_o$ with an apparent half-maximum inhibitory $[Ca^{2+}]$ of $IC_{50} \approx 220$ μM and a Hill coefficient of 2.1 (ref. 2). Extracellular Ca^{2+} inhibited ATP release with $IC_{50} = 495$ μM and a Hill coefficient of 1.9 (Fig. 3e, f). CALHM1 ion currents are also activated by membrane depolarization[2]. At normal levels of $[Ca^{2+}]_o$, hCALHM1-expressing, but not mock-transfected, cells released ATP in response to high-$[K^+]_o$-induced depolarization (Fig. 3g, h). Depolarization-induced ATP release was inhibited by RuR but not by heptanol or probenecid (Fig. 3i, j), which are respectively connexin and pannexin 1 blockers[17]. Expression of mouse CALHM1 conferred ATP release with similar properties (Supplementary Fig. 8). Regulation of ATP release in hCALHM1-expressing cells is therefore similar to that of CALHM1 channel gating, indicating that the CALHM1 channel is a conduit for ATP release.

The three types of taste bud cell can be classified on the basis of whole-cell electrophysiological fingerprints[13,18]. We verified these fingerprints by recording whole-cell currents in single isolated taste bud cells from TRPM5–green fluorescent protein (GFP) mice with GFP expressed specifically in type II cells (Supplementary Fig. 9). With tetraethylammonium in the extracellular solution and Cs^+ as the major cation in the pipette solution[2] (to block K^+ currents), the electrophysiological subtypes were identified by the combination of voltage-gated Na^+ currents (I_{Na}) and non-selective, slowly activating outward currents (I_{slow}) with inward tail currents at −70 mV (I_{tail}) (Fig. 4a–c and Supplementary Fig. 9). To identify CALHM1 ion currents, we recorded whole-cell currents in isolated taste bud cells from wild-type and *Calhm1*$^{-/-}$ mice (Fig 4a–c). Loss of CALHM1 substantially reduced the amplitudes of I_{slow} and I_{tail} without affecting the amplitude of I_{Na} in type II cells (Fig. 4d–f), whereas no differences were observed in type I and type III cells (Fig. 4b, c, g). The slowly activating outward current in type II cells was inhibited by the non-specific CALHM1 blocker Gd^{3+} but not by probenecid or heptanol, ruling out contributions of pannexin 1 or connexins (Fig. 4h). These data demonstrate that CALHM1 channel conductance is present in type II cells and contributes the major component of the slowly activating outward current. Notably, the amount of depolarization-induced

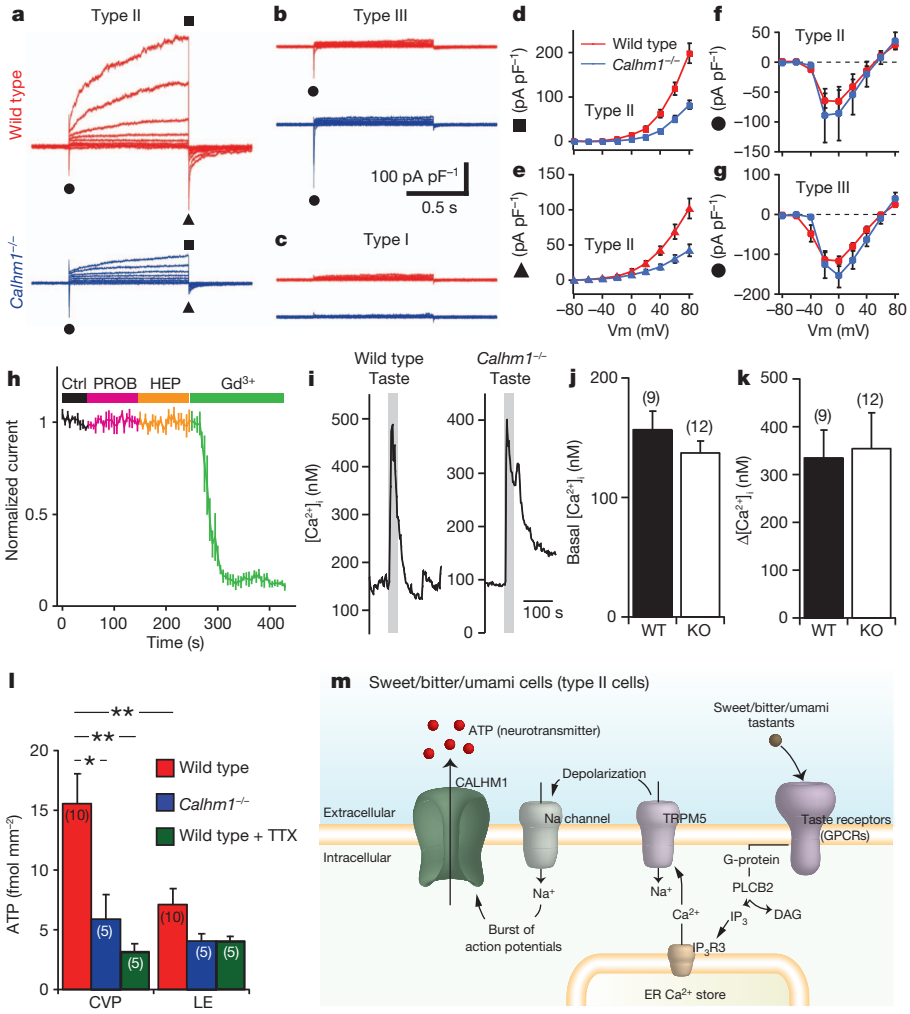


Figure 4 | CALHM1 is required for taste-evoked ATP release from taste cells.
a–c, Electrophysiological phenotypes of type I, type II and type III cells identified in wild-type (red) and *Calhm1*$^{-/-}$ (blue) taste cells. Cells held at −70 mV and pulsed from −80 to +80 mV in 20-mV increments with 1-s duration. Type I current in **c** recorded from 16 wild-type and 9 *Calhm1*$^{-/-}$ cells. **d–g**, I_{Na} (circle; **f**), I_{slow} at end of pulses (square; **d**) and I_{tail} (triangle; **e**) measured for type II cells ($n = 9$ wild type, $n = 10$ *Calhm1*$^{-/-}$), and I_{Na} measured for type III cells ($n = 9$ wild type, $n = 6$ *Calhm1*$^{-/-}$) (**g**). Vm, membrane voltage. **h**, Sensitivities of I_{slow} in GFP-positive cells from TRPM5–GFP mice to Gd^{3+} (100 μM), probenecid (1 mM), 1-heptanol (1 mM) ($n = 4$). **i**, $[Ca^{2+}]_i$ in type II cells from wild-type (left, 9 cells) and *Calhm1*$^{-/-}$ (right, 12 cells) mice. Type II cells identified by robust $[Ca^{2+}]_i$ response to a mix of sweet and bitter substances (grey bar). **j**, **k**, Basal (**j**) and taste-evoked (**k**) responses are comparable in wild-type and *Calhm1*$^{-/-}$ cells. KO, knockout; WT, wild type. **l**, Taste-evoked ATP release from gustatory CVP tissue and non-gustatory lingual epithelium. Bitter mix elicits considerable ATP release from CVP versus lingual epithelium in wild-type mice; this is abolished in *Calhm1*$^{-/-}$ mice and by 1 μM tetrodotoxin (TTX). Error bars, s.e.; $*P < 0.05$, $**P < 0.01$ (Student's *t*-test). **m**, Schematic illustration of signal-transduction cascade in type II taste receptor cells with CALHM1 as the ATP-release pathway.

©2013 Macmillan Publishers Limited. All rights reserved

ATP release from type II cells is correlated with the magnitude of I_{slow} (ref. 13).

Type II cells detect sweet, bitter and umami compounds via G-protein-coupled taste receptors[4] that stimulate a common signal-transduction cascade involving activation of PLCB2, inositol 1,4,5-trisphosphate-mediated Ca^{2+} release and Ca^{2+}-dependent activation of TRPM5 channels[11] that depolarizes the plasma membrane to generate action potentials and subsequent non-vesicular release of ATP. Importantly, no differences in basal $[Ca^{2+}]_i$ and taste-evoked $[Ca^{2+}]_i$ responses were observed between wild-type and *Calhm1*$^{-/-}$ type II cells (Fig. 4i–k). Furthermore, *Calhm1* deficiency had no effect on I_{Na} (Fig. 4f), TRPM5 expression in type II cells (Fig. 1a and Supplementary Figs 1i and 2) or the ATP content of taste buds (Supplementary Fig. 10). Strikingly, however, taste-evoked release of ATP from circumvallate papillae was abolished in *Calhm1*$^{-/-}$ mice (Fig. 4l and Supplementary Fig. 11). This strongly suggests that CALHM1 contributes to the major ATP-release mechanism in sweet/bitter/umami-sensing taste bud cells (Fig. 4l, m).

Our study sheds light on a novel cellular ATP-releasing mechanism by demonstrating that CALHM1 is a voltage-gated ion channel that mediates tetrodotoxin-sensitive ATP release in taste buds (Fig. 4l) as an essential mechanism of sweet, bitter and umami taste perception. As such, CALHM1 provides a missing link in the signal-transduction cascade in type II cells, connecting taste receptor activation and the generation of Na^+ action potentials to the activation of afferent gustatory neural pathways[1] (Fig. 4m). It has not escaped our attention that other CALHM isoforms[6] as well as pannexin 1 and connexins are also present in taste buds[9,12,13] and might also be involved in ATP release in taste, perhaps acting in parallel with or in a complex with CALHM1. Signalling by extracellular ATP is a widespread phenomenon that regulates many physiological activities[19], including neurotransmission[20,21], intercellular communication[22,23], vascular tone[24] and sensory transduction[5,25–27]. CALHM1 and its isoforms may participate in physiologically important ATP release elsewhere.

METHODS SUMMARY

All animal experiments were conducted according to protocols approved by the Institutional Animal Care and Use Committee of the University of Pennsylvania, the Feinstein Institute for Medical Research, the Monell Chemical Senses Center and the University of Minnesota Duluth.

Full Methods and any associated references are available in the online version of the paper.

Received 6 September 2012; accepted 15 January 2013.
Published online 6 March 2013.

1. Chaudhari, N. & Roper, S. D. The cell biology of taste. *J. Cell Biol.* **190,** 285–296 (2010).
2. Ma, Z. *et al.* Calcium homeostasis modulator 1 (CALHM1) is the pore-forming subunit of an ion channel that mediates extracellular Ca^{2+} regulation of neuronal excitability. *Proc. Natl Acad. Sci. USA* **109,** E1963–E1971 (2012).
3. Siebert, A. P. *et al.* Structural and functional similarities of calcium homeostasis modulator 1 (CALHM1) ion channel with connexins, pannexins and innexins. *J. Biol. Chem.* **288,** 6140–6153 (2013).
4. Chandrashekar, J., Hoon, M. A., Ryba, N. J. & Zuker, C. S. The receptors and cells for mammalian taste. *Nature* **444,** 288–294 (2006).
5. Finger, T. E. *et al.* ATP signaling is crucial for communication from taste buds to gustatory nerves. *Science* **310,** 1495–1499 (2005).
6. Dreses-Werringloer, U. *et al.* A polymorphism in CALHM1 influences Ca^{2+} homeostasis, Aβ levels, and Alzheimer's disease risk. *Cell* **133,** 1149–1161 (2008).
7. Koppel, J. *et al.* CALHM1 P86L polymorphism modulates CSF Aβ levels in cognitively healthy individuals at risk for Alzheimer's disease. *Mol. Med.* **17,** 974–979 (2011).
8. Lambert, J. C. *et al.* The CALHM1 P86L polymorphism is a genetic modifier of age at onset in Alzheimer's disease: a meta-analysis study. *J. Alzheimers Dis.* **22,** 247–255 (2010).
9. Moyer, B. D. *et al.* Expression of genes encoding multi-transmembrane proteins in specific primate taste cell populations. *PLoS ONE* **4,** e7682 (2009).
10. Matsumoto, I., Ohmoto, M., Narukawa, M., Yoshihara, Y. & Abe, K. Skn-1a (Pou2f3) specifies taste receptor cell lineage. *Nature Neurosci.* **14,** 685–687 (2011).
11. Zhang, Y. *et al.* Coding of sweet, bitter, and umami tastes: different receptor cells sharing similar signaling pathways. *Cell* **112,** 293–301 (2003).
12. Huang, Y. J. *et al.* The role of pannexin 1 hemichannels in ATP release and cell-cell communication in mouse taste buds. *Proc. Natl Acad. Sci. USA* **104,** 6436–6441 (2007).
13. Romanov, R. A. *et al.* Afferent neurotransmission mediated by hemichannels in mammalian taste cells. *EMBO J.* **26,** 657–667 (2007).
14. Vandenbeuch, A., Zorec, R. & Kinnamon, S. C. Capacitance measurements of regulated exocytosis in mouse taste cells. *J. Neurosci.* **30,** 14695–14701 (2010).
15. Romanov, R. A. *et al.* Dispensable ATP permeability of pannexin 1 channels in a heterologous system and in mammalian taste cells. *J. Cell Sci.* **125,** 5514–5523 (2012).
16. Sabirov, R. Z. & Okada, Y. ATP release via anion channels. *Purinergic Signal.* **1,** 311–328 (2005).
17. Silverman, W., Locovei, S. & Dahl, G. Probenecid, a gout remedy, inhibits pannexin 1 channels. *Am. J. Physiol. Cell Physiol.* **295,** C761–C767 (2008).
18. Romanov, R. A. & Kolesnikov, S. S. Electrophysiologically identified subpopulations of taste bud cells. *Neurosci. Lett.* **395,** 249–254 (2006).
19. Burnstock, G. & Kennedy, C. P2X receptors in health and disease. *Adv. Pharmacol.* **61,** 333–372 (2011).
20. Edwards, F. A., Gibb, A. J. & Colquhoun, D. ATP receptor-mediated synaptic currents in the central nervous system. *Nature* **359,** 144–147 (1992).
21. Evans, R. J., Derkach, V. & Surprenant, A. ATP mediates fast synaptic transmission in mammalian neurons. *Nature* **357,** 503–505 (1992).
22. Cotrina, M. L., Lin, J. H., Lopez-Garcia, J. C., Naus, C. C. & Nedergaard, M. ATP-mediated glia signaling. *J. Neurosci.* **20,** 2835–2844 (2000).
23. Cotrina, M. L., Lin, J. H. & Nedergaard, M. Cytoskeletal assembly and ATP release regulate astrocytic calcium signaling. *J. Neurosci.* **18,** 8794–8804 (1998).
24. Burnstock, G. Dual control of local blood flow by purines. *Ann. NY Acad. Sci.* **603,** 31–44 (1990).
25. Chizh, B. A. & Illes, P. P2X receptors and nociception. *Pharmacol. Rev.* **53,** 553–568 (2001).
26. Cockayne, D. A. *et al.* Urinary bladder hyporeflexia and reduced pain-related behaviour in P2X3-deficient mice. *Nature* **407,** 1011–1015 (2000).
27. Wynn, G., Rong, W., Xiang, Z. & Burnstock, G. Purinergic mechanisms contribute to mechanosensory transduction in the rat colorectum. *Gastroenterology* **125,** 1398–1409 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by a KeySpan award to P.M., several US NIH grants (GM56328, MH059937, NS072775 to J.K.F.; DC10393 to M.G.T.; EY13624 to M.M.C.; R03DC011143 to I.M.; Core Grant P30 EY001583 to the University of Pennsylvania; Core Grant P30DC011735 to the Monell Chemical Senses Center) and the University of Minnesota's Undergraduate Research Opportunities Program to S.L. and M.A. A.T. and M.O. are JSPS Fellows. We thank R. F. Margolskee for the TRPM5–GFP mice and Y. Ninomiya for comments on the manuscript.

Author Contributions A.T., V.V., A.L., Z.M., M.O., I.M., H.Z., L.A., S.L., M.A., G.H., G.D. and N.C. designed and performed experiments. P.M. and V.V. generated the *Calhm1* knockout mice. J.K. and P.D. designed experiments. M.G.T., M.M.C., P.M. and J.K.F. designed experiments and helped with data interpretation. A.T., J.K.F. and P.M. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.M. (pmaramba@nshs.edu) and J.K.F. (foskett@mail.med.upenn.edu).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

***Calhm1*$^{-/-}$ mice.** *Calhm1*$^{+/-}$ founder mice were generated at genOway (Lyon, France) as described previously[2]. Briefly, *Calhm1* exon 1 deletion was performed by homologous recombination in 129Sv embryonic stem cells. Positive clones were screened by PCR and confirmed by Southern blot analysis. The resultant embryonic stem cells were injected into blastocysts derived from C57BL/6J mice to obtain chimaeric mice that possessed germline transmission of the targeted *Calhm1* locus. Mice on the mixed 129Sv × C57BL/6J genetic background were used in this study. Wild-type (+/+) and *Calhm1* knockout (−/−) mice of F_1 obtained from littermate mice were used for two-bottle preference tests. All other experiments were performed with wild-type and knockout littermates. In some mice, the neomycin cassette was removed and the resulting strain was backcrossed with C57BL/6J mice. Briefly, *Calhm1*$^{+/-}$ mice were mated with mice bearing an *EIIa–cre* transgene (B6.FVB-Tg (EIIa-cre)C_{5379}Lmgd/J, The Jackson Laboratory) on the C57BL/6J background to remove the neomycin resistance cassette by *cre*–loxP-mediated excision. The resulting *Calhm1*$^{+/-}$ pups were further backcrossed to C57BL/6J (The Jackson Laboratory) for at least five generations before being made homozygous. The removal of the neomycin cassette was confirmed by PCR. Backcrossed mice were used for two-bottle preference tests. Loss of CALHM1 expression in *Calhm1*$^{-/-}$ taste buds was verified by reverse-transcription (RT)–PCR and *in situ* hybridization.

Laser capture microdissection, RNA amplification and RT–PCR. Circumvallate taste tissue from *Calhm1*$^{+/+}$ and *Calhm1*$^{-/-}$ mice was embedded in cryomolds using OCT freezing medium. Twelve micrometre-thick tissue sections were cut on a Leica CM1850 cryostat (Leica Microsystems), collected on RNase-free membrane slides (Leica) and stained with cresyl violet. Taste buds from the CVP and surrounding lingual epithelium were isolated using a Leica laser microdissection system LMD7000 (Leica). Taste buds and surrounding epithelium were pooled from a total of four mice of each genotype. Total RNA from taste bud and lingual epithelium samples was purified using a RNAqueous-Micro Kit (Ambion). Total RNA was amplified using two sequential rounds with MessageAmp II aRNA kit (Ambion), as per the manufacturer's instructions. One microgram of RNA was transcribed into complementary DNA (cDNA) using Invitrogen's SuperScript III First-Strand Synthesis System for RT–PCR with random hexamers, according to the supplied protocol. A 50-µl PCR reaction was run with the following final concentrations: 450 ng of each primer (see Supplementary Table 4 for PCR primer sequences), 2 mM $MgCl_2$, 0.3 mM dNTPs, 2.5 U Taq polymerase (Promega GoFlex DNA polymerase) and 1 µl of 10 ng µl^{-1} DNA. PCR cycling conditions used for *Calhm1*: 95 °C for 3 min; 35 cycles of 95 °C for 30 s, 65 °C for 30 s and 72 °C for 45 s; 72 °C for 7 min; 4 °C (hold).

***In situ* hybridization.** Oral epithelia containing taste buds were dissected from adult male mice deeply anaesthetized with an overdose of sodium pentobarbital (Abbott Laboratories) and embedded in the frozen OCT compound (Sakura Finetech USA). Fresh-frozen sections of 8-µm thickness were prepared using a cryostat (CM3050S, Leica Microsystems). The *in situ* hybridization procedure was described previously[28]. In brief, digoxigenin- and fluorescein-labelled antisense RNAs prepared using RNA labelling mix (Roche Diagnostics) and an RNA polymerase (Stratagene) were used for hybridization after fragmentation under alkaline conditions to a length of about 150 bases. Fresh-frozen sections were fixed with 4% PFA, treated with diethylpyrocarbonate, prehybridized with salmon sperm DNA for 2 h at 65 °C and hybridized with antisense riboprobes for 40 h at 65 °C. For single labelling of *Tas2r108* and *Trpm5*, however, prehybridization and hybridization were carried out at 58 °C. After hybridization, the sections were washed in ×0.2 SSC at 58 °C or 65 °C and blocked with 0.5% blocking reagent (Roche Diagnostics) in Tris-buffered saline. Chromogenic signals, except those from *Calhm1*, were developed for one day using alkaline phosphatase-conjugated anti-digoxigenin antibody (1:500, Roche Diagnostics) and 4-nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indolyl-phosphate as chromogenic substrates. Chromogenic *Calhm1* signals were developed for two days to clarify the cells expressing *Calhm1* messenger RNA. No *Calhm1* signal was observed in *Skn-1* or *Calhm1* knockout taste tissues after signal development for three days. Stained images were obtained with a Nikon Eclipse 80i microscope (Nikon Instruments Inc.) equipped with a DXM1200C digital camera (Nikon). Fluorescent signals were developed using biotin-conjugated anti-fluorescein antibody (1:500, Vector Laboratories), avidin–biotin complex (Vector Laboratories), tyramide signal amplification biotin system (1:50, PerkinElmer) and Alexa 488-conjugated streptavidin (4 µg ml^{-1}, Invitrogen). Fluorescent images were obtained with a Leica SP2 confocal scanning microscope (Leica). For double labelling of *Calhm1* with *Trpm5* and *Tas1r3*, fluorescent signals were first developed and obtained, then chromogenic signals were developed and obtained as described above, and finally the images were superimposed using Photoshop CS3 (Adobe). RNA probes generated were to nucleotides 1–894 of *Tas2r108* (GenBank accession number, AF227148), nucleotides 310–3,491 of *Trpm5* (accession number, AF228681), nucleotides 525–2,725 of *Tas1r3* (accession number, AF337039), and nucleotides 1–1,407 and 2,148–2,369 of *Calhm1* cDNA fragment, which contains 1,047 base pairs of the entire coding sequence and 1,322 base pairs of 3′-non-coding region.

Immunofluorescence staining. For fixed-tissue preparation, mice 12 weeks old or older were anaesthetized with an overdose of sodium pentobarbital (Abbott Laboratories) and transcardially perfused with ice-cold PBS followed by 4% paraformaldehyde in ice-cold PBS. Tongue epithelia containing CVP were dissected and postfixed with the same fixative at 4 °C overnight. Tissue samples were then cryoprotected in 30% sucrose in PBS, embedded in frozen OCT compound (Sakura Finetech) and sectioned using a cryostat (CM1900, Leica Microsystems) at 8 µm. Tissue sections were mounted on tissue-adhesive-coated glass slides (Fisher Scientific) and stored at −80 °C until analysed. For immunofluorescence staining, slides were rinsed with PBS, incubated in a preheated target retrieval solution (S1700, Dako) at 80 °C for 20 min, allowed to cool to room temperature (20–25 °C) for 20 min, still in the target retrieval solution, and washed in PBS for 3 × 10 min. After blocking by 1-h incubation in PBS containing 5% normal donkey serum (PBS-5% NDS) at room temperature, sections were incubated overnight at 4 °C with primary antibodies diluted in PBS-5% NDS: 1:3,000 for rabbit anti-TRPM5 (Alomone Labs); 1:500 for rabbit anti-PLCB2 (Santa Cruz Biotechnology); 1:500 for rabbit anti-AADC (Gene Tex); 1:500 for goat anti-KCNQ1 (Santa Cruz). The next day, the slides were washed for 3 × 10 min in PBS at room temperature and incubated for 1 h with Alexa-Fluor-conjugated antibodies diluted in PBS-5% NDS: 1:500 for Alexa Fluor 488 donkey anti-rabbit IgG; 1:500 for Alexa Fluor 568 anti-goat IgG (Invitrogen). Finally, slides were washed for 3 × 10 min in PBS at room temperature and mounted in VectaShield with 4′,6-diamidino-2-phenylindole (H-1500, Vector Laboratories). Stained sections were imaged with a confocal scanning microscope (LSM 710, Carl Zeiss).

Single-taste cell analyses. Circumvallate taste epithelium was enzymatically delaminated, taste buds were collected from peeled epithelium and dissociated single-taste cells were collected, all as detailed previously[29]. For patch clamp and fura-2 Ca^{2+} imaging experiments, isolated cells were placed on poly-L-lysine-coated coverslips and allowed to settle for 30–60 min.

Total RNAs isolated from taste buds and from adjacent non-taste epithelium were used as positive and negative controls, respectively. Individual dissociated taste cells from PLCB2–GFP (type II) and GAD1–GFP (type III) transgenic mice were selected for collection on the basis of their expression of GFP, as previously described[30]. Type I taste cells were selected from PLCB2–GFP × GAD1–GFP double-transgenic mice by the absence of GFP expression[30]. Total RNA isolated from taste cells was either subjected to T7 RNA polymerase-based linear amplification (aRNA, Message BOOSTER kit, Epicentre) or was directly converted to cDNA, both as previously described[29]. For aRNA-based analyses, 1% of the resulting cDNA was used as template for PCR. For direct RT–PCR, 15% of cDNA from each cell was used to test for diagnostic messenger RNAs and 30% was used for *Calhm1*. PCR conditions with HotStarTaq Plus (QIAGEN) were 95 °C for 5 min followed by 45 90-s cycles, each of which comprised three 30-s components, one at 94 °C, one at the annealing temperature and one at 72 °C. PCR primers, annealing temperatures and product sizes were as in Supplementary Table 4. Two different primer sets were used for *Calhm1* (Supplementary Table 4). Primers 1 and 2 were used for RT–PCR on bulk tissue RNA, as well as for aRNA from individual taste cells (Fig. 1a and Supplementary Fig. 3a, b, d). Primers 3 and 4 span an intron, and were used for the single-cell direct RT–PCR to avoid amplifying from genomic DNA (Supplementary Fig. 3c). The cell type of all single cells and pools of cells was confirmed by RT–PCR for three diagnostic cell-type markers: *Entpd2* (type I), *Plcb2* (type II) and *Snap25* (type III).

Two-bottle preference tests. *Calhm1*$^{-/-}$ and wild-type control mice were presented for 48 h with two drinking bottles, one containing water and the other a solution supplemented with ascending concentrations of the specific taste compounds to be tested. All solutions were prepared with tap water and served at room temperature. Mice had *ad libitum* access to a standard chow diet (Laboratory Rodent Diet 5001, LabDiet, PMI Nutrition International) and drinking solutions. Two independent groups of mice were tested for the following series of taste solutions. The order of testing was as follows: sucrose, saccharin, NaCl, denatonium (group 1); MSG, quinine, HCl (group 2). Sucrose, saccharin, denatonium benzoate, quinine sulphate and MSG were from Sigma-Aldrich. HCl and NaCl were from Thermo Fisher Scientific Inc. The order of stimulus presentation was identical for wild-type and knockout mice. The mice had three days with a single bottle of drinking water between each test series. The two bottles were switched after 24 h to control for a possible effect of the bottle positions. The change in liquid levels was considered to be the mouse's liquid intake[31]. Fluid intakes were measured volumetrically to the nearest 0.1 ml. The volumes of water and taste compound solutions consumed were recorded. Solution preference was calculated as the intake of taste solution divided by total liquid intake, and this ratio was expressed as a percentage. The taste compounds were chosen as exemplars of the

©2013 Macmillan Publishers Limited. All rights reserved

sweet, bitter, umami, salty and sour taste qualities, and their concentrations were chosen to span the range between indifference and marked acceptance (for sucrose, saccharin and MSG) or avoidance (for quinine sulphate, denatonium benzoate, NaCl and HCl). Results were analysed by mixed-design analyses of variance with factors of group (wild type or *Calhm1*$^{-/-}$) and concentration. Differences between the groups in consumption of specific concentrations of taste solution were determined using Bonferroni *post hoc* tests. Differences in response of each group to individual concentrations of test compounds relative to water were determined using Student's *t*-test. Results presented in Fig. 2a are representative examples taken from the more comprehensive concentration functions displayed in Supplementary Fig. 4a and Supplementary Tables 1 and 2.

Brief-exposure gustometer tests. Acceptance of taste solution was assessed with brief-exposure tests using procedures similar to those used by other groups[11,32–35].

Twelve *Calhm1*$^{-/-}$ mice (5 male, 7 female) and 12 wild-type littermate controls (5 male, 7 female) were maintained in a vivarium at 23 °C on a 12 h:12 h light/dark cycle with lights off at 19:00. When not being tested, each mouse was housed alone in a plastic 'tub' cage (26.5 cm × 17 cm × 12 cm) with a stainless-steel grid lid and wood shavings scattered on the floor. The mouse ate pelleted AIN-76A diet and drank deionized water according to the regimen described below. Throughout the experiment, each mouse was weighed daily, immediately before it was placed into a gustometer.

Brief-exposure taste acceptance was assessed using three MS160-Mouse gustometers manufactured by DiLog Instruments. Each gustometer consists of a 14.5 cm × 30 cm × 15 cm test chamber with a motorized shutter that controls access to a taste solution. Bottles of taste solution are mounted on a rack that is precisely positioned by a stepper motor so that any one of eight different taste solutions can be presented to the mouse. The drinking spout of each bottle is part of a high-frequency alternating current contact circuit so that each lick the mouse makes is detected and recorded. Details of construction and other technical information is available elsewhere[34,36]. Taste compounds were reagent grade and were purchased from Sigma-Aldrich and dissolved in room-temperature deionized water. Deionized water was used as the water (0 concentration) and for 'washout' trials (see below). To avoid any undue influence of subtle differences among the three gustometers we used, we ensured that each mouse was always tested with the same gustometer and that equal numbers of wild-type and *Calhm1*$^{-/-}$ mice were tested with each gustometer.

To train the mice to sample taste solutions, they were first deprived of water for 22.5 h and then placed in a gustometer with its shutter open. During this first training session, each mouse had continuous access to water for 25 min from the time it first licked the drinking spout. It was then returned to its home cage and given water for 1 h. During the following two days, this procedure was repeated, except the shutter allowing access to water was closed 5 s after each time the mouse began to lick, and was re-opened after a 7.5-s interval. Once again, after 25 min, the mouse was returned to its home cage and given water for 1 h. By the second test using these procedures, all except one wild-type and one *Calhm1*$^{-/-}$ mouse had learned to obtain water during the 5-s access periods. We continued to test the two non-responders and they began to drink during the third or fifth test session (see below).

The mice received two test sessions with each of the 11 compounds presented in Supplementary Fig. 4b. These were presented in two cycles, with exemplars of the four basic tastes in the first cycle (sucrose, quinine hydrochloride, HCl and NaCl) and the other compounds in the second cycle. Only one taste compound was used during a session (see below) and there was only one session per day. The deprivation regimen used to investigate the response to sucrose, saccharin and Polycose differed from that used to investigate the response to the other taste compounds because these are hedonically positive and the other three compounds are hedonically negative[34]. Prior to a session with a hedonically positive compound, each mouse received free access to food and water for 24 h. It then received 1 g of food and 2 ml of water, and the session began 24 h later. After these sessions, the mouse had a recovery day with free access to food and water for 24 h. Its water was then removed for 22.5 h to prepare it for the next session. After sessions with hedonically negative compounds, each mouse received water for 1 h in its home cage and it was then deprived of water in preparation for the next session. All sessions lasted 25 min. When hedonically positive compounds were being investigated, the session began with a single test of the highest concentration available to kindle the mouse's interest in the drinking spout. After this, repeated series of five concentrations (including water) were presented, in a quasi-random order (a concentration could appear only once in a series of five tests). For each exposure, the shutter was open for 5 s, during which licks of the drinking spout were counted. This was followed by 7.5 s with the shutter closed, during which a new taste solution was positioned, ready for the next presentation. For sessions with hedonically negative compounds, repeated series of five or six concentrations (including water) were presented in randomized order. Additional 1-s washout trials with water were interposed between test trials. Thus, a mouse received access to a taste solution for 5 s followed by 7.5 s with the shutter closed, then access to water for 1 s followed by 7.5 s with the shutter closed, followed by the next taste solution for 5 s, and so on. We think the brief washout trials with water have the effect of cleansing the mouse's palate and thus prevent it from quitting because it expects only bad-tasting solutions.

Separate analyses were conducted for each taste compound. The mean number of licks by each mouse in response to each concentration was obtained by averaging the results of identical exposures. These values for individual mice were then used in mixed-design analyses of variance with factors of group (wild type or *Calhm1*$^{-/-}$) and concentration. Mice that did not respond during any presentation of a particular concentration of a taste compound were not included in statistical analyses of that compound. *Post hoc* least-significant difference tests were used to assess between-group differences in consumption of specific concentrations of taste solution and differences in the response of each group to individual concentrations of each taste compound (STATISTICA 10, Stat Soft Inc.). All analyses were conducted using a criterion for significance of $P < 0.05$. Results presented in Fig. 2b are representative examples taken from the more comprehensive concentration functions displayed in Supplementary Fig. 4b and Supplementary Table 3.

Whole-chorda-tympani nerve recordings. The mice were anaesthetized with an intramuscular injection of a mixture of 1.75 mg ml^{-1} ketamine and 1.75 mg ml^{-1} xylazine in saline (5 µl g^{-1} body weight). Anaesthesia was maintained with 0.4–0.6% isoflurane. Body temperature, surgical table temperature, blood oxygen, anaesthesia level and heart rate were continuously monitored. The chorda tympani nerve was dissected free from its junction with the lingual nerve to the tympanic bulla, where the central part of the chorda tympani was cut and the peripheral part then mobilized in the rostral direction so that afferent nerve impulses could be recorded. The nerve impulses were recorded between a silver wire electrode and an indifferent electrode touching the walls of the wound, fed into a custom-made amplifier and monitored over a loudspeaker and an oscilloscope. The nerve impulses were processed by a smoothed absolute value circuit integrator[37] and rectified to a d.c. potential whose amplitude was related to the nerve impulse frequency, here called the summated signal. This signal and a code related to the kind of taste compound used were fed into an oscilloscope and computer. Both the individual nerve impulses and the summated signal were recorded (Gould ES 1000). The stimuli were delivered to the tongue with an open flow system controlled by the computer under conditions of constant flow and temperature[37] (33 °C). Each stimulation lasted for 5 s with 50-s rinsing time between stimulations. Care was exercised to make sure that the flow was directed over the fungiform papillae. In each animal, we strived to run the same sequence of stimuli at least three times while we recorded the nerve impulses from the whole chorda tympani nerve. The integrated response during stimulation was calculated by subtracting the area of nerve activity preceding stimulation from that during stimulation. Thus, the data reflect the level of activity during stimulation time. The responses to all compounds were expressed relative to the response to 0.1 M NH_4Cl in each mouse. The average in each animal and group was calculated, the variance determined and the significance between the chorda tympani responses of *Calhm1*$^{-/-}$ and wild-type mice determined. $P < 0.05$ was considered significant.

Cell culture. HeLa cells (American Type Culture Collection) were grown in plastic flasks at 37 °C in a humidified incubator in culture medium containing 90% (v/v) DMEM, 10% fetal bovine serum and ×1 Antibiotic-Antimycotic (Invitrogen) with 5% CO_2-in-air.

ATP measurements using cultured cells. Extracellular ATP concentration was measured by the luciferin–luciferase reaction as previously described[38]. HeLa cells were seeded onto 96-well microplates (Corning Costar) at a density of 10,000 cells per well one day before transfection. Cells in each well were transfected with 0.2 µg of hCALHM1 cDNA, mouse CALHM1 cDNA or empty vector (pIRES2.EGFP) using 0.4 µl Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. After 20–30 h, media were removed and cells were incubated for 1 h before 75 µl of the 100-µl bath solution containing 2 mM Ca^{2+} was replaced with an equal volume of test solution to establish final $[Ca^{2+}]_o$ and drug concentrations. The plate was immediately placed in a microplate luminometer (Synergy 2, BioTEK) and 10 µl of ATP assay solution (FL-AAM and FL-AAB, Sigma-Aldrich) was dispensed into each well. ATP release was measured every 2 min. ATP concentration was calculated from a standard curve created in each plate. Separate standard curves were made in experiments involving all added drugs at the concentrations used. The bath solution contained (in mM) 150 NaCl, 5 KCl, 10 HEPES and 10 glucose, pH 7.4, adjusted with NaOH. $[Ca^{2+}]_o$ values below 90 µM were achieved by mixing $CaCl_2$ and EGTA, and free $[Ca^{2+}]_o$ was calculated using WEBMAXC software (http://maxchelator.stanford.edu). When a drug was tested, cells were pre-incubated for 1 h with the drug before being exposed to low-

©2013 Macmillan Publishers Limited. All rights reserved

$[Ca^{2+}]_o$ solution containing the same drug. The high-$[K^+]_o$ condition (117.5 mM) was established by replacing Na^+ with equimolar K^+.

ATP measurements using intact taste buds. Taste-evoked ATP release from intact taste buds was recorded with sheets of lingual epithelium as described previously[5]. The tongue epithelium was cut into pieces containing the CVP or lingual epithelium devoid of taste buds and mounted in a custom Ussing-type chamber that separated the fluid-containing serosal chamber from the apical surface. The apical surface was selectively exposed to a bitter mix (40 mM denatonium benzoate, 1 mM cycloheximide, 10 mM HEPES, pH 7.4) for 45 s. The serosal fluid (130 µl) was then collected and added to an equal volume of ATP assay solution (FL-AAM and FL-AAB, Sigma-Aldrich) to determine its ATP content by the luciferin–luciferase assay. The solution on the serosal side contained (in mM) 150 NaCl, 5 KCl, 1.5 $CaCl_2$, 1 $MgCl_2$ 10 HEPES and 10 glucose, pH 7.4, adjusted with NaOH.

Single-cell Ca^{2+} imaging. Isolated taste cells on coverslips were secured in a perfusion chamber and mounted on the stage of an inverted microscope (Nikon Eclipse TE2000). Cells were loaded with fura-2AM (Molecular Probes; 2.5 µM) for 45 min at room temperature in the bath solution containing (in mM) 150 NaCl, 5 KCl, 2 $CaCl_2$, 1 $MgCl_2$, 10 HEPES and 10 glucose, pH 7.4, adjusted with NaOH. Cells were continuously perfused with the bath solution and stimulated for 30 s by bath perfusion of taste mix (2 mM saccharin, 100 µM SC45647, 1 mM denatonium, 10 µM cycloheximide; dissolved in the bath solution at pH 7.4). Fura-2 was alternately illuminated at 340 or 380 nm, and the emitted fluorescence intensity at 510 nm was collected with a PerkinElmer Ultraview imaging system. The background fluorescence was estimated for a region without cells and subtracted during analysis. Changes in $[Ca^{2+}]_i$ are presented as changes in fluorescence ratio (F340/F380). Dye calibration was achieved by applying experimentally determined constants to the equation $[Ca^{2+}] = K_d\beta(R - R_{min})/(R_{max} - R)$. Macros used for analysis were custom macros written for IGOR PRO (WaveMetrics). Isolated taste bud cells that responded to a taste stimulus by robust $[Ca^{2+}]_i$ increase were identified as type II cells: 24% of wild-type cells (9 out of 38 cells from three animals) and 39% of *Calhm1*$^{-/-}$ cells (12 out of 31 cells from four animals).

Single-cell electrophysiology. Whole-cell recordings were made in single taste bud cells isolated from *Calhm1*$^{-/-}$ and wild-type littermates and TRPM5–GFP mice with a patch-clamp amplifier (Axopatch 200B, Axon Instruments). TRPM5–GFP mice were a gift from Dr R. F. Margolskee[39] and were used for selective recordings from type II cells on the basis of their expression of GFP. The pipette solution contained (in mM) 155 CsF, 2 $MgCl_2$, 1 $CaCl_2$, 11 EGTA, 10 HEPES, pH 7.3, adjusted with methanesulphonic acid, ~308 mOsm. The bath solution contained (in mM) 140 Na^+, 5.4 K^+, 10 TEA^+, 20 sucrose, 1.5 Ca^{2+}, 1 Mg^{2+}, 6 Cl^- and 10 HEPES, pH 7.4, adjusted with methanesulphonic acid, ~330 mOsm. Patch pipettes were fabricated from thick-walled borosilicate glass capillaries (PG10150-4, World Precision Instruments) and had a resistance of 5–9 MΩ in the recording condition. Electrode capacitance was compensated electronically, and 70% of series resistance (10–25 MΩ) was compensated with a lag of 20 µs. Currents were low-pass-filtered at 5 kHz with eight-pole Bessel characteristics and sampled at 10 kHz. For testing pharmacological properties of I_{slow} in type II cells (Fig. 4h), whole-cell currents were recorded in GFP-positive taste cells isolated from TRPM5–GFP mice and I_{slow} was evoked every 5 s by a voltage pulse to +100 mV with 1-s duration from a holding potential of −70 mV. Current amplitudes were taken at the end of the pulse and normalized to the average of values before drug addition (initial 50 s) for each cell.

28. Ohmoto, M., Matsumoto, I., Misaka, T. & Abe, K. Taste receptor cells express voltage-dependent potassium channels in a cell age-specific manner. *Chem. Senses* **31,** 739–746 (2006).
29. Dvoryanchikov, G., Sinclair, M. S., Perea-Martinez, I., Wang, T. & Chaudhari, N. Inward rectifier channel, ROMK, is localized to the apical tips of glial-like cells in mouse taste buds. *J. Comp. Neurol.* **517,** 1–14 (2009).
30. Sinclair, M. S. *et al.* Oxytocin signaling in mouse taste buds. *PLoS ONE* **5,** e11980 (2010).
31. Tordoff, M. G. & Bachmanov, A. A. Mouse taste preference tests: why only two bottles? *Chem. Senses* **28,** 315–324 (2003).
32. Damak, S. *et al.* Trpm5 null mice respond to bitter, sweet, and umami compounds. *Chem. Senses* **31,** 253–264 (2006).
33. Eddy, M. C. *et al.* Double P2X2/P2X3 purinergic receptor knockout mice do not taste NaCl or the artificial sweetener SC45647. *Chem. Senses* **34,** 789–797 (2009).
34. Glendinning, J. I., Gresack, J. & Spector, A. C. A high-throughput screening procedure for identifying mice with aberrant taste and oromotor function. *Chem. Senses* **27,** 461–474 (2002).
35. Hallock, R. M., Tatangelo, M., Barrows, J. & Finger, T. E. Residual chemosensory capabilities in double P2X2/P2X3 purinergic receptor null mice: intraoral or postingestive detection? *Chem. Senses* **34,** 799–808 (2009).
36. Spector, A. C., Andrews-Labenski, J. & Letterio, F. C. A new gustometer for psychophysical taste testing in the rat. *Physiol. Behav.* **47,** 795–803 (1990).
37. Hellekant, G. & Roberts, T. W. in *Experimental Cell Biology of Taste and Olfaction: Current Techniques and Protocols* (eds Spielman, A. I. & Brand, J. G.) 277–290 (CRC, 1995).
38. Li, A., Leung, C. T., Peterson-Yantorno, K., Mitchell, C. H. & Civan, M. M. Pathways for ATP release by bovine ciliary epithelial cells, the initial step in purinergic regulation of aqueous humor inflow. *Am. J. Physiol. Cell Physiol.* **299,** C1308–C1317 (2010).
39. Clapp, T. R., Medler, K. F., Damak, S., Margolskee, R. F. & Kinnamon, S. C. Mouse taste cells with G protein-coupled taste receptors lack voltage-gated calcium channels and SNAP-25. *BMC Biol.* **4,** 7 (2006).

©2013 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature11926

CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance

Adam Greenbaum[1]*, Yen-Michael S. Hsu[2]*, Ryan B. Day[1], Laura G. Schuettpelz[3], Matthew J. Christopher[1], Joshua N. Borgerding[1], Takashi Nagasawa[4,5] & Daniel C. Link[1]

Haematopoietic stem cells (HSCs) primarily reside in the bone marrow where signals generated by stromal cells regulate their self-renewal, proliferation and trafficking. Endosteal osteoblasts[1,2] and perivascular stromal cells including endothelial cells[3], CXCL12-abundant reticular cells[4,5], leptin-receptor-positive stromal cells[6], and nestin–green fluorescent protein (GFP)-positive mesenchymal progenitors[7] have all been implicated in HSC maintenance. However, it is unclear whether specific haematopoietic progenitor cell (HPC) subsets reside in distinct niches defined by the surrounding stromal cells and the regulatory molecules they produce. CXCL12 (chemokine (C–X–C motif) ligand 12) regulates both HSCs and lymphoid progenitors and is expressed by all of these stromal cell populations[7–11]. Here we selectively deleted *Cxcl12* from candidate niche stromal cell populations and characterized the effect on HPCs. Deletion of *Cxcl12* from mineralizing osteoblasts has no effect on HSCs or lymphoid progenitors. Deletion of *Cxcl12* from osterix-expressing stromal cells, which include CXCL12-abundant reticular cells and osteoblasts, results in constitutive HPC mobilization and a loss of B-lymphoid progenitors, but HSC function is normal. *Cxcl12* deletion from endothelial cells results in a modest loss of long-term repopulating activity. Strikingly, deletion of C*xcl12* from nestin-negative mesenchymal progenitors using *Prx1–cre* (*Prx1* also known as *Prrx1*) is associated with a marked loss of HSCs, long-term repopulating activity, HSC quiescence and common lymphoid progenitors. These data suggest that osterix-expressing stromal cells comprise a distinct niche that supports B-lymphoid progenitors and retains HPCs in the bone marrow, and that expression of CXCL12 from stromal cells in the perivascular region, including endothelial cells and mesenchymal progenitors, supports HSCs.

CXCL12 has a crucial role in maintaining HSC function, including retention in the bone marrow[8,12–14], quiescence[15,16] and repopulating activity[16]. To test the hypothesis that CXCL12 production by different stromal cell populations has distinct effects on HSCs and lineage-committed HPCs, we generated a floxed allele of *Cxcl12* (*Cxcl12*^{fl}) to delete *Cxcl12* conditionally from candidate niche cells in the bone marrow (Supplementary Fig. 2). Deletion of *Cxcl12* from endothelial cells and mature osteoblasts was mediated by the *Tie2–cre* recombinase and osteocalcin (*Oc*)–*cre* transgenes, respectively (*Oc* also known as *Bglap*). To target *Cxcl12* deletion from osteoprogenitors, we used the osterix (*Osx*)–*cre* transgene (*Osx* also known as *Sp7*), which mediates efficient recombination in mature osteoblasts and osteoblast progenitors[17]. It also targets CXCL12-abundant reticular (CAR) cells, a perivascular stromal cell population implicated in HSC and B-lymphoid progenitor (BLP) maintenance[5,11]. Finally, we used the *Prx1–cre* transgene to target multipotent mesenchymal progenitors in the appendicular skeleton. PRX1 is a transcription factor expressed early during limb bud mesoderm development, and *Prx1–cre* targets all cells derived from limb bud mesoderm[18]. Lineage mapping studies were performed using a *Cxcl12*^{gfp} knock-in mouse to define CAR cells[11]. These studies showed that both the *Osx–cre* and *Prx1–cre* transgenes efficiently targeted recombination in mature osteoblasts, osteocytes and CAR cells in long bones (Fig. 1a–d and Supplementary Fig. 3).

We generated triple-transgenic mice containing one floxed *Cxcl12* allele, one null allele (*Cxcl12*$^{fl/-}$) and a *cre*-recombinase transgene. Total CXCL12 messenger RNA (mRNA) expression in the femoral bone marrow of *Oc–cre*- and *Tie2–cre*-targeted mice was similar to that observed in control mice (Fig. 1e). In contrast, CXCL12 mRNA expression was reduced by 70% in *Osx–cre*-targeted mice and was nearly undetectable in *Prx1–cre*-targeted mice. A similar decrease in CXCL12 protein levels was observed (Fig. 1f). To confirm *Cxcl12* deletion from CAR cells, mice containing *Cxcl12*$^{fl/gfp}$ and either the *Osx–cre* or the *Prx1–cre* transgene were generated (the *Cxcl12*^{gfp} allele is a null allele). Indeed, CXCL12 mRNA, although still detectable, was markedly reduced in CAR cells that were sorted from these mice (Fig. 1g). As expected, CXCL12 mRNA was nearly undetectable in endothelial cells sorted from *Tie2–cre*-targeted mice (Fig. 1h). Together these data suggest that, under basal conditions, the majority of CXCL12 is produced by CAR cells, with mature osteoblasts and endothelial cells being only minor contributors.

All conditional knockout mice had normal peripheral blood counts and the same relative percentage of granulocytes, monocytes, B cells and T cells (Supplementary Table 1). However, bone marrow cellularity in femurs was reduced by approximately 50% in both the *Osx–cre*- and *Prx1–cre*-targeted mice, which was due, in part, to a loss of B cells. HPC subsets in the bone marrow were quantified by flow cytometry (Fig. 2a). The number of c-KIT$^+$SCA1$^+$LIN$^-$ (KSL) cells, short-term HSCs, multipotent progenitors and myeloid-committed progenitors was similar in all mice with the exception of a twofold decrease in common myeloid progenitors in *Prx1–cre*-targeted mice (Supplementary Fig. 4). Loss of CXCL12 expression in endothelial cells or mature osteoblasts had no effect on the number of phenotypic HSCs (Fig. 2b–d). The frequency of phenotypic HSCs in the bone marrow of *Osx–cre*-targeted mice was comparable to control mice (data not shown); however, because bone marrow cellularity was reduced, a modest decrease in the absolute number of HSCs was observed. In contrast, a significant decrease in both the frequency and absolute number of phenotypic HSCs in *Prx1–cre*-targeted mice was observed, with nearly undetectable dormant HSCs (FLK2$^-$CD34$^-$CD150$^+$CD48$^-$ KSL cells). Consistent with these findings, competitive repopulation assays showed a significant multilineage long-term repopulating defect using bone marrow from *Prx1–cre*- but not *Osx–cre*-targeted mice (Fig. 3a, b). Despite the normal number of phenotypic HSCs, a small but significant decrease in long-term repopulating activity also was observed using *Tie2–cre*-targeted bone marrow. Serial transplantation of bone marrow from *Prx1–cre*- or *Tie2–cre*-targeted mice showed no further decrease in repopulating activity in secondary

[1]Department of Medicine, Washington University School of Medicine, St Louis, Missouri 63110, USA. [2]Department of Pathology & Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. [3]Department of Pediatrics, Washington University School of Medicine, St Louis, Missouri 63110, USA. [4]Department of Immunobiology and Hematology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto 606-8507, Japan. [5]Japan Science and Technology Agency, Core Research for Evolutionary Science and Technology, Tokyo 102-0075, Japan.
*These authors contributed equally to this work.

©2013 Macmillan Publishers Limited. All rights reserved

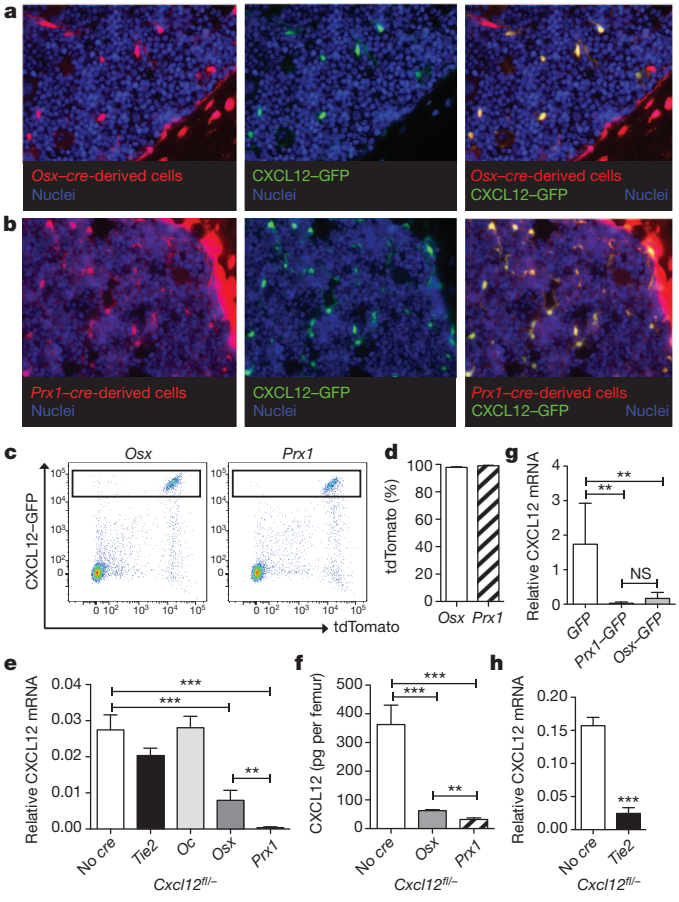


Figure 1 | Targeting *Cxcl12* deletion in bone marrow stromal cell populations. **a**, **b**, Lineage mapping was by performed by generating *Osx–cre ROSA26*$^{Ai9/+}$ *Cxcl12*$^{gfp/+}$ mice (**a**) or *Prx1–cre ROSA26*$^{Ai9/+}$ *Cxcl12*$^{gfp/+}$ mice (**b**). Cells that had undergone *cre*-mediated recombination express tdTomato (red). Cells that express CXCL12 also express GFP (green). Counterstaining with 4′,6-diamidino-2-phenylindole (DAPI) highlights nuclei (blue). Shown are representative photomicrographs of femur sections. The left panel shows tdTomato fluorescence, the middle panel shows GFP fluorescence and the right panel shows both. Original magnification, ×40. **c**, Representative dot plots showing GFP and tdTomato expression in CD45$^-$LIN$^-$ stromal cells harvested from *Osx–cre ROSA*$^{Ai9/+}$ *Cxcl12*$^{gfp/+}$ mice (left) or *Prx1–cre ROSA26*$^{Ai9/+}$ *Cxcl12*$^{gfp/+}$ mice (right). GFP^{bright} CAR cells are boxed. **d**, Percentage of CAR cells that express tdTomato ($n = 5$). **e**, CXCL12 mRNA expression relative to β-actin mRNA, shown on total bone marrow RNA ($n = 6–13$). **f**, CXCL12 protein in bone marrow extracellular fluid as measured by enzyme-linked immunosorbent assay ($n = 3–5$). **g**, CAR cells were sorted from *Osx–cre Cxcl12*$^{gfp/fl}$ (*Osx–GFP*) or *Prx1–cre Cxcl12*$^{gfp/fl}$ (*Prx1–GFP*) mice, and RNA was prepared. Shown is CXCL12 mRNA expression relative to β-actin mRNA ($n = 5–7$). **h**, CD31$^+$LIN$^-$CD45$^-$ endothelial cells were sorted from *Tie2–cre*-targeted and control mice, and RNA was prepared. Shown is CXCL12 mRNA expression relative to β-actin mRNA ($n = 3$). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. NS, not significant.

recipients, suggesting that self-renewal capacity may be restored when HSCs are exposed to a normal stromal microenvironment (Supplementary Fig. 5). Quiescence is a fundamental property of HSCs and is closely related to long-term repopulating activity[19]. Increased cycling of HSCs was observed in *Prx1–cre*- but not *Osx–cre*- or *Tie2–cre*-targeted cells. In contrast, increased cycling of more mature KSL progenitors was observed in both *Prx1–cre*- and *Osx–cre*-targeted cells (Fig. 3c, d). Collectively, these data show that CXCL12 production by *Prx1–cre*-targeted stromal cells and, to a lesser extent, endothelial cells is required for maintenance of HSC repopulating activity and quiescence. Consistent with results from a companion paper[20], our data suggest that CXCL12 production by mature osteoblasts and osteoblast precursors is dispensable for HSC maintenance.

Because CXCL12 has been shown to be important in the retention of HPCs within the bone marrow[5,21], we next quantified HPCs in the blood and spleen. In *Osx–cre*-targeted mice, the number of colony-forming cells and KSL cells was increased in the blood and spleen, demonstrating constitutive HPC mobilization (Fig. 3e, f and Supplementary Fig. 6). Notably, although CXCL12 expression in the bone marrow is significantly lower in *Prx1–cre*-targeted mice relative to *Osx–cre*-targeted mice, both mouse lines showed a similar magnitude of HPC mobilization. Thus, our data suggest that, although CXCL12 production from *Osx–cre*-targeted stromal cells is largely dispensable for HSC maintenance, it is required for the efficient retention of HPCs in the bone marrow.

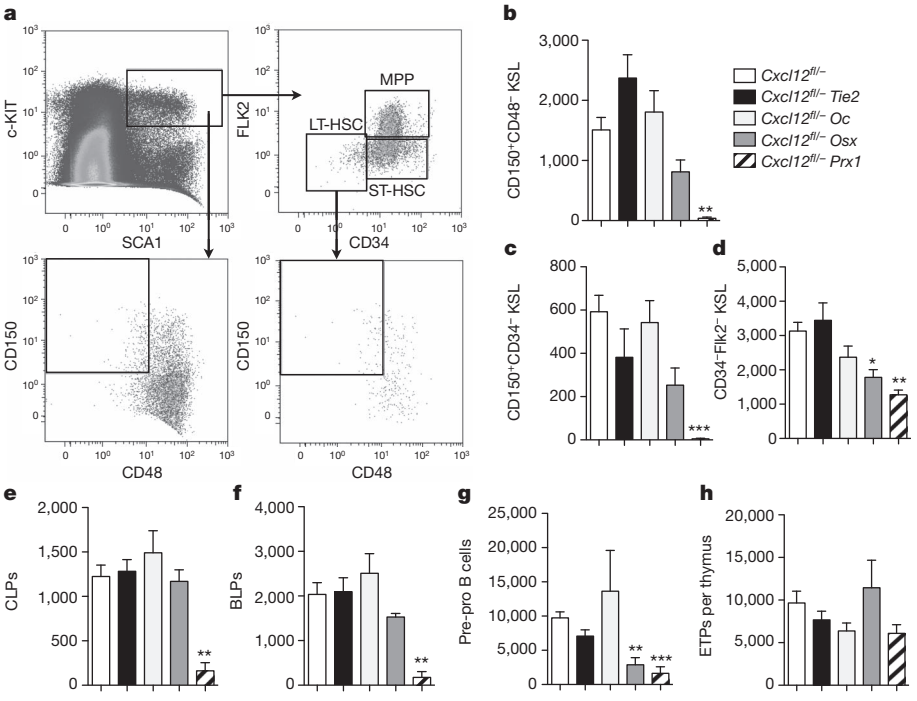


Figure 2 | Deletion of *Cxcl12* in defined stromal cell population results in the selective loss of HSCs and lymphoid progenitors. **a**, Representative dot plots showing the gating strategy to identify HPC populations. Data are gated on lineage$^-$ cells. MPP, multipotent progenitor cell; LT-HSC, long-term HSC; ST-HSC, short-term HSC. **b**–**g**, Numbers of CD150$^+$CD48$^-$CD41$^-$ KSL cells (**b**), CD150$^+$CD41$^-$CD48$^-$CD34$^-$FLK2$^-$ KSL cells (dormant HSCs) (**c**), CD34$^-$FLK2$^-$ KSL cells (**d**), common lymphoid progenitors (CLPs) (**e**), BLPs (**f**) and pre-pro B cells (**g**) per femur. **h**, The number of earliest thymic progenitors (ETPs) per thymus. Data represent the mean ± s.e.m. of 6–13 mice. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

©2013 Macmillan Publishers Limited. All rights reserved

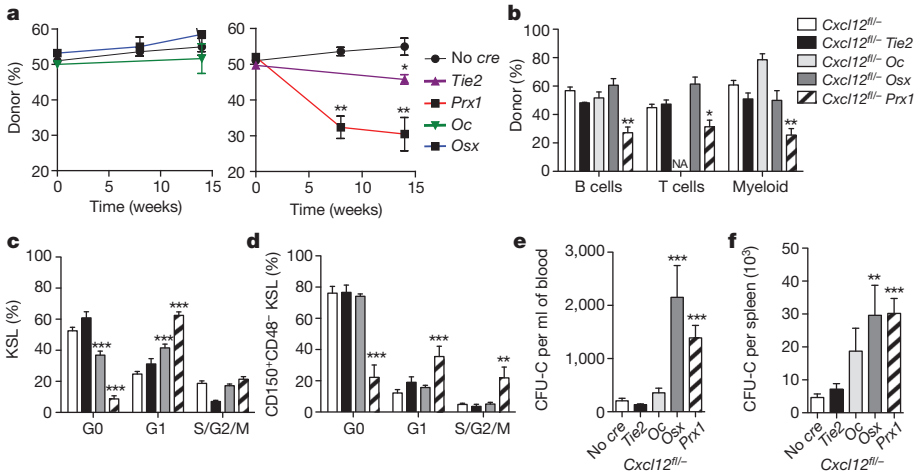


Figure 3 | Deletion of *Cxcl12* in defined stromal cell populations results in HPC mobilization and a selective loss of repopulating activity and HSC quiescence. **a**, Competitive repopulation assays were performed with a 1:1 ratio of donor and wild-type competitor bone marrow. Shown is the percentage of donor peripheral leukocytes over time. Data represent the mean ± s.e.m. of 10–16 mice. **b**, Percentage of donor cells in the bone marrow of the indicated lineage 14 weeks after transplantation. Data represent the mean ± s.e.m. of 5 or 6 mice. **c**, **d**, Percentage of cells in the indicated stage of the cell cycle for KSL (**c**) and $CD150^+CD48^-CD41^-$ KSL (**d**) cells. **e**, **f**, Numbers of colony-forming cells (CFU-C) in the blood (**e**) and spleen (**f**). Data represent the mean ± s.e.m. of 5–15 mice. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. NA, not available.

CXCL12 is required for normal B- and T-cell development[12,22]. Pre-pro B cells are found in close association with CAR cells[11], and ablation of CAR cells is associated with a loss of common lymphoid progenitors (CLPs) and pro B cells[4]. Here we show that deletion of *Cxcl12* from mineralizing osteoblasts or endothelial cells has no effect on CLPs, BLPs or pre-pro B cells (Fig. 2e–h and Supplementary Fig. 7). In contrast, *Cxcl12* deletion from CAR cells using *Osx–cre* results in a marked loss of pre-pro B cells and a trend towards a loss of BLPs. However, the number of CLPs and earliest thymic progenitors in the thymus are normal. Deletion of *Cxcl12* by *Prx1–cre*-targeted stromal cells results in a similar phenotype but also results in a marked loss of CLPs. Deletion of *Cxcl12* in osteoblasts using *Col2.3–cre* also results in a modest decrease in CLPs and lymphoid-primed multipotent progenitors[20]. Together, these data suggest that CXCL12 production by CAR cells or osteoblast precursors, but not mineralizing osteoblasts or endothelial cells, is required for the maintenance of B-lymphoid-committed progenitors, and that CLP maintenance is supported by CXCL12 production from both endosteal osteoblasts and a *Prx1–cre*-targeted perivascular stromal cell population. The normal CLPs in *Osx–cre*-targeted mice may be secondary to compensatory changes related to the severe loss of pre-pro B cells.

We next performed studies to define the stromal cell populations differentially targeted by *Prx1–cre* and *Osx–cre*. We first considered the possibility that *Prx1–cre* may target endothelial cells in the bone marrow. However, we detected no tdTomato expression in bone marrow endothelial cells from *Prx1–cre* reporter mice (Fig. 4a, b). Moreover, expression of CXCL12 mRNA from sorted $CD31^+$ endothelial cells from *Prx1–cre*-targeted mice was comparable to that in control mice (Supplementary Fig. 8). Thus, loss of CXCL12 from bone marrow endothelial cells does not account for the loss of HSCs in *Prx1–cre*-targeted mice.

We extended the lineage mapping studies to the $CD45^-$ $lineage^-$ $PDGFR\alpha^+SCA1^+$ (PαS) cell population, which is enriched for mesenchymal stem cells[23]. Whereas *Osx–cre* did not target this cell population, approximately 50% of cells were targeted by *Prx1–cre*, including a subpopulation that expressed intermediate levels of CXCL12 (Fig. 4c–e). To evaluate the mesenchymal progenitor activity of the *Prx1–cre*-targeted cells, we sorted *Prx1–cre*-targeted ($tdTomato^+$) and non-targeted

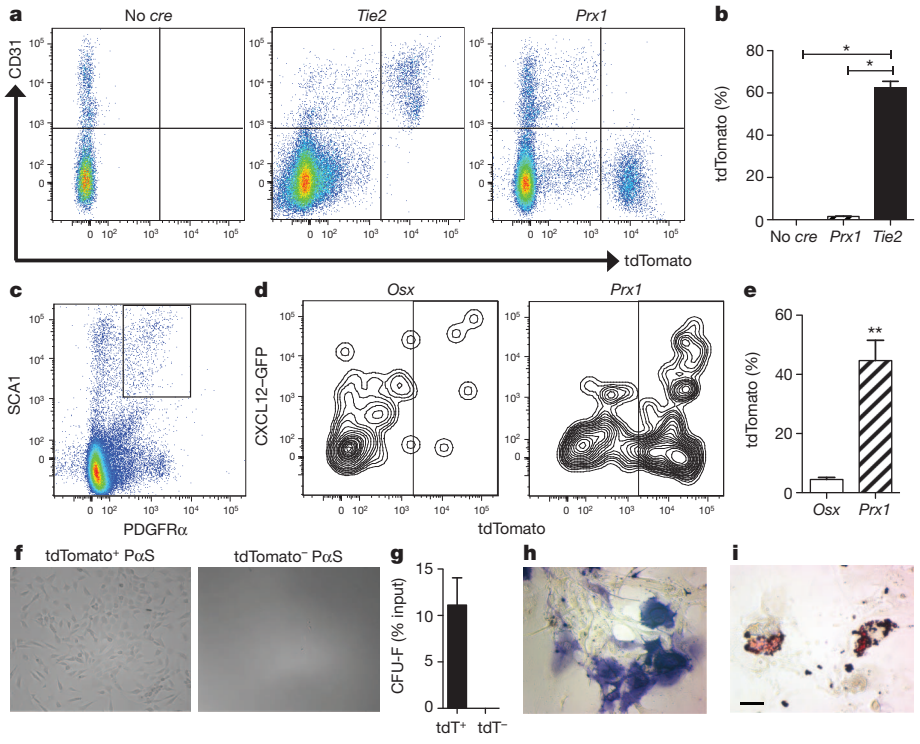


Figure 4 | *Prx1–cre* differentially targets a $PDGFR\alpha^+SCA1^+$ CXCL12-expressing mesenchymal progenitor cell population. **a**, Representative dot plots showing tdTomato and CD31 expression in bone marrow cells from the indicated mice. Data are gated on $CD45^-LIN^-$ cells. **b**, Percentage of $CD31^+CD45^-LIN^-$ cells that express tdTomato. (**c**) Representative dot plot showing the gating strategy to identify $CD45^-LIN^-PDGFR\alpha^+SCA1^+$ cells (boxed). Data are gated on $CD45^-LIN^-$ cells. **d**, Representative dot plots showing GFP and tdTomato expression in $CD45^-LIN^-$ PαS cells from *Osx–cre ROSA*$^{Ai9/+}$ *Cxcl12*$^{gfp/+}$ mice (left panel) and *Prx1–cre ROSA26*$^{Ai9/+}$ *Cxcl12*$^{gfp/+}$ mice (right panel). **e**, Percentage of LIN^- PαS cells that express tdTomato ($n = 3$). $*P < 0.001$, $**P < 0.01$. **f**, $tdTomato^+$ and $tdTomato^-$ PαS cells were sorted from *Prx1–cre ROSA26*$^{Ai9/+}$ *Cxcl12*$^{gfp/+}$ mice and CFU-F assays were performed. Shown are representative photomicrographs of the cultures on day 14. **g**, Number of CFU-Fs per 100 sorted PαS cells. Data represent the mean ± s.e.m. of three independent experiments. **h**, **i**, Cells were harvested from the CFU-F cultures on day 14 and replated under osteogenic (**h**) or adipogenic (**i**) culture conditions. Shown are representative photomicrographs of cells stained for alkaline phosphatase (**h**) and oil red O (**i**). Scale bar, 50 μm.

©2013 Macmillan Publishers Limited. All rights reserved

PαS cells and assessed their clonogenic capacity. All of the colony-forming unit fibroblast (CFU-F) activity was contained within the *Prx1*-targeted PαS cell population, with more than 10% of these cells having CFU-F activity (Fig. 4f, g). This compares with a CFU-F frequency of approximately 4% in unselected PαS cells[23] and less than 1% in nestin–GFP^{+} stromal cells[7]. The *Prx1*-targeted PαS cells have osteogenic and adipogenic differentiation potential *in vitro*, consistent with a mesenchymal stem cell phenotype (Fig. 4h, i). RNA expression profiling of *Prx1*-targeted PαS cells is notable for the lack of nestin[7], CD146 (ref. 9) and leptin receptor[6], all of which have been used to mark stromal cells contributing to HSC maintenance (Supplementary Table 2). Notably, other than moderate CXCL12 expression, these cells do not express genes classically associated with HSC maintenance, including kit ligand[6] and angiopoietin-1 (ref. 19), although high expression of several matrix proteins implicated in HPC regulation (for example proteoglycan 4 (ref. 24) and osteonectin[25]) is present.

Collectively, these data suggest that distinct stromal cell niches in the bone marrow exist that regulate specific HPC populations (Supplementary Fig. 1). Osterix-expressing stromal cells comprise a niche that supports B-lymphoid progenitors and retains HPCs in the bone marrow, whereas CXCL12 production from nestin^{-} leptin receptor^{-} mesenchymal progenitors is required for HSC and CLP maintenance. However, our results do not rule out the possibility that CXCL12 expression from a small subset of CAR cells also contributes to HSC and CLP maintenance.

METHODS SUMMARY

Targeting of the *Cxcl12* locus was accomplished by conventional techniques. Conditional knockout mice were produced by interbreeding *Cxcl12*$^{fl/fl}$ mice with *Osx–cre* (ref. 26), *Prx1–cre* (ref. 18), *Oc–cre* (ref. 27), *Tie2–cre* (ref. 28) or *Cxcl12*$^{+/-}$ mice[22]. Lineage mapping was accomplished using Ai9 (ref. 29) and *Cxcl12*^{gfp} (ref. 12) mice. All mice with the exception of *Cxcl12*^{gfp} mice were maintained on the C56Bl6/J background. Unless indicated otherwise, data are presented as mean ± s.e.m. and were analysed using Student's *t*-test, one-way analysis of variance or two-way analysis of variance.

Full Methods and any associated references are available in the online version of the paper.

Received 7 August 2012; accepted 21 January 2013.
Published online 24 February 2013.

1. Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425,** 841–846 (2003).
2. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425,** 836–841 (2003).
3. Hooper, A. T. *et al.* Engraftment and reconstitution of hematopoiesis is dependent on VEGFR2-mediated regeneration of sinusoidal endothelial cells. *Cell Stem Cell* **4,** 263–274 (2009).
4. Omatsu, Y. *et al.* The essential functions of adipo-osteogenic progenitors as the hematopoietic stem and progenitor cell niche. *Immunity* **33,** 387–399 (2010).
5. Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25,** 977–988 (2006).
6. Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481,** 457–462 (2012).
7. Méndez-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466,** 829–834 (2010).
8. Peled, A. *et al.* Dependence of human stem cell engraftment and repopulation of NOD/SCID mice on CXCR4. *Science* **283,** 845–848 (1999).
9. Sacchetti, B. *et al.* Self-renewing osteoprogenitors in bone marrow sinusoids can organize a hematopoietic microenvironment. *Cell* **131,** 324–336 (2007).
10. Semerad, C. L. *et al.* G-CSF potently inhibits osteoblast activity and CXCL12 mRNA expression in the bone marrow. *Blood* **106,** 3020–3027 (2005).
11. Tokoyoda, K., Egawa, T., Sugiyama, T., Choi, B. I. & Nagasawa, T. Cellular niches controlling B lymphocyte behavior within bone marrow during development. *Immunity* **20,** 707–718 (2004).
12. Ara, T. *et al.* A role of CXC chemokine ligand 12/stromal cell-derived factor-1/pre-B cell growth stimulating factor and its receptor CXCR4 in fetal and adult T cell development *in vivo*. *J. Immunol.* **170,** 4649–4655 (2003).
13. Bonig, H., Priestley, G. V., Nilsson, L. M., Jiang, Y. & Papayannopoulou, T. PTX-sensitive signals in bone marrow homing of fetal and adult hematopoietic progenitor cells. *Blood* **104,** 2299–2306 (2004).
14. Kawabata, K. *et al.* A cell-autonomous requirement for CXCR4 in long-term lymphoid and myeloid reconstitution. *Proc. Natl Acad. Sci. USA* **96,** 5663–5667 (1999).
15. Nie, Y., Han, Y. C. & Zou, Y. R. CXCR4 is required for the quiescence of primitive hematopoietic cells. *J. Exp. Med.* **205,** 777–783 (2008).
16. Tzeng, Y. S. *et al.* Loss of Cxcl12/Sdf-1 in adult mice decreases the quiescent state of hematopoietic stem/progenitor cells and alters the pattern of hematopoietic regeneration after myelosuppression. *Blood* **117,** 429–439 (2011).
17. Maes, C. *et al.* Osteoblast precursors, but not mature osteoblasts, move into developing and fractured bones along with invading blood vessels. *Dev. Cell* **19,** 329–344 (2010).
18. Logan, M. *et al.* Expression of Cre recombinase in the developing mouse limb bud driven by a Prxl enhancer. *Genesis* **33,** 77–80 (2002).
19. Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118,** 149–161 (2004).
20. Ding, L. & Morrison, S. J. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature* http://dx.doi.org/10.1038/nature11885 (2013).
21. Foudi, A. *et al.* Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nature Biotechnol.* **27,** 84–90 (2009).
22. Nagasawa, T. *et al.* Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature* **382,** 635–638 (1996).
23. Morikawa, S. *et al.* Prospective identification, isolation, and systemic transplantation of multipotent mesenchymal stem cells in murine bone marrow. *J. Exp. Med.* **206,** 2483–2496 (2009).
24. Zhan, M. & Han, Z. C. Hemangiopoietin inhibits apoptosis of MO7e leukemia cells through phosphatidylinositol 3-kinase-AKT pathway. *Biochem. Biophys. Res. Commun.* **317,** 198–204 (2004).
25. Lehmann, S. *et al.* Common deleted genes in the 5q- syndrome: thrombocytopenia and reduced erythroid colony formation in SPARC null mice. *Leukemia* **21,** 1931–1936 (2007).
26. Rodda, S. J. & McMahon, A. P. Distinct roles for Hedgehog and canonical Wnt signaling in specification, differentiation and maintenance of osteoblast progenitors. *Development* **133,** 3231–3244 (2006).
27. Zhang, M. *et al.* Osteoblast-specific knockout of the insulin-like growth factor (IGF) receptor gene reveals an essential role of IGF signaling in bone matrix mineralization. *J. Biol. Chem.* **277,** 44005–44012 (2002).
28. Kisanuki, Y. Y. *et al.* Tie2-Cre transgenic mice: a new model for endothelial cell-lineage analysis *in vivo*. *Dev. Biol.* **230,** 230–242 (2001).
29. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13,** 133–140 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Woloszynek, F. Liu, A. Khalaf and M. Romine for technical assistance; G. Callis, S. Oh and M. Vig for technical advice; J. Tucker-Davis for animal care; and T. Clemens for the *Oc–cre* mice. This work was supported by NIH grants RO1 HL60772 (D.C.L.) and F30 HL097423 (A.G.).

Author Contributions A.G. and Y.-M.S.H. designed and performed the research, analysed the data and wrote the manuscript. L.G.S., J.N.B. and R.B.D. performed experiments characterizing haematopoiesis in the conditional *Cxcl12*-deficient mice. M.J.C. designed and cloned the CXCL12 conditional knockout construct. D.C.L. supervised all of the research and edited the manuscript, which was approved by all co-authors.

Author Information The RNA expression profiling data have been deposited in Gene Expression Omnibus under accession number GSE43613. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.C.L. (dlink@dom.wustl.edu).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Mice. With the exception of *Cxcl12*^{gfp} mice, all transgenic strains had been backcrossed at least ten generations onto a C57BL/6 background. *Osx–cre* (ref. 26), *Prx1–cre* (ref. 18), *Tie2–cre* (ref. 28), and Ai9 (B6.Cg-Gt(ROSA)26Sor$^{tm9(CAG-tdTomato)Hze}$/J) (ref. 29) mice were obtained from The Jackson Laboratory. *EIIa–cre* mice[30] were a gift from Monica Bessler (University of Pennsylvania). *Oc–cre* mice[27] were a gift from Thomas Clemens (Johns Hopkins University, Maryland) and *Cxcl12*^{gfp} mice[12] were a gift from Takashi Nagasawa (Kyoto University, Japan). *Cxcl12*$^{+/-}$ mice[22] were obtained through the RIKEN BioResource Center (Ibaraki, Japan). Mice were maintained under standard pathogen-free conditions according to methods approved by the Washington University Animal Studies Committee.

Generation of *Cxcl12*^{fl} mice. A floxed allele of *Cxcl12* was generated containing *LoxP* sites flanking exon 2 of *Cxcl12*; a third *LoxP* was inserted 3′ of the neomycin selection cassette (Supplementary Fig. 2). Generation of targeted embryonic stem cells and blastocyst injections were performed as previously described[31]. Excision of the neomycin cassette was accomplished through partial recombination by intercrossing mice with mice expressing *EIIa–cre*. Mice were genotyped using PCR primers: *Cxcl12*^{flox} forward, 5′-CTACACCTCCTCTAGGTAAACCAGTCA GCC-3′; *Cxcl12*^{flox} reverse, 5′-GGACACCAGAACCTTGAAACTGACA-3′.

Bone marrow transplantation. Bone marrow from wild-type *Ly5.1/5.2*-expressing mice was mixed in a 1:1 ratio with bone marrow from experimental or control mice expressing the *Ly5.2* locus. A total of 2×10^6 cells were injected retro-orbitally into lethally irradiated (1,000 cGy), wild-type *Ly5.1*-expressing mice. Because *Prx1* is expressed primarily in limb-bud-derived long bones, only tibias and femurs were used for transplant and other analyses.

Blood, bone marrow, spleen and thymus analysis. Blood, bone marrow and spleen cells and thymocytes were collected using standard techniques and quantified using a Hemavet automated cell counter (CDC Technologies).

Flow cytometry. Cells were stained by standard protocols with the following antibodies (eBiosciences unless otherwise noted). Lineage analysis and chimaerism was assessed using peridinin chlorophyll protein complex (PerCP)-Cy5.5-conjugated Ly5.1 (A20, CD45.1), allophycocyanin (APC)-conjugated Ly5.2 (104, CD45.2) and one or more of the following lineage markers: APC-conjugated CD115 (AFS98, monocytes), fluorescein isothiocyanate (FITC)-conjugated Ly6C/G (RB6-8C5, Gr-1, myeloid), phycoerythrin (PE)-conjugated CD3e (145-2C11, T lymphocytes), and APC-eFluor780-conjugated CD45R (RA3-6B2, B220, B lymphocytes).

For haematopoietic stem and progenitor cell analysis, cells were stained with a cocktail of biotin-conjugated B220, TER119, CD3e, GR1 and CD41 (MWReg30); PE-conjugated CD150 (TC15-12F12.2, Biolegend); PE-Cy7-conjugated CD48 (HM48-1, BD Biosciences); PerCP-Cy5.5-conjugated SCA1 (D7); APC eFluor780-conjugated c-Kit (2B-8); FITC-conjugated CD34 (RAM34); APC-conjugated FLK2 (A2F10); eFluor450-conjugated CD16/32; and eFluor605NC-conjugated streptavidin. For HSC cell cycle staining, cells were stained with the biotin-conjugated lineage panel, PE-conjugated CD150, PE-Cy7-conjugated CD48, PerCP-Cy5.5-conjugated SCA1, APC-conjugated c-KIT and APC-eFluor780-conjugated streptavidin. Cells were then fixed using the Cytofix/Cytoperm kit (BD Biosciences), stained with FITC-conjugated Ki-67 (B56, BD Biosciences) and resuspended in 1 mg ml^{-1} of DAPI. Doublets were gated out using FSC versus FSC-W. Data were collected on a Gallios 10-colour, 3-laser flow cytometer (Beckman Coulter). Data were analysed with FLOWJO (Treestar).

For CLP/BLP analysis, bone marrow cells were stained with a cocktail of PE-Cy7-conjugated B220, TER119, CD3e and GR-1; APC-conjugated CD27 (LG.7F9); biotinylated IL7Rα (gift from Deepta Bhattacharya, Washington University); PE-conjugated FLK2; FITC-conjugated Ly6D (49-H4, BD Biosciences); and eFluor450-conjugated streptavidin. CLPs were defined as B220$^-$TER119$^-$ CD3e$^-$GR-1$^-$CD27$^+$IL7Rα$^+$FLK2$^+$Ly6D$^-$ cells and BLPs were defined as B220$^-$TER119$^-$CD3e$^-$GR-1$^-$CD27$^+$IL7Rα$^+$FLK2$^+$Ly6D$^+$ cells.

For analysis of pre-pro B cells, bone marrow cells were stained with APC-eFluor780-conjugated B220, a cocktail of PerCP-Cy5.5-conjugated CD3e, CD11c (N418) and NK1.1 (PK136), APC-conjugated IgM (II/4), eFluor450-conjugated IgD (11-26c), PE-Cy7-conjugated CD19 (eBio1D3), PE-conjugated CD43 (S7, BD Biosciences) and FITC-conjugated Ly6D (BD Biosciences). Pre-pro B cells were defined as B220$^+$CD3e$^-$CD11c$^-$NK1.1$^-$IgM$^-$IgD$^-$CD19$^-$CD43$^+$Ly6D$^+$ cells.

For analysis of earliest thymic progenitors, thymocytes were stained using a cocktail of FITC-conjugated CD4 (RM4-5), CD8 (53-6.7) and CD11b (M1/70); a cocktail of PE-Cy7-conjugated B220, TER119, CD3e and GR-1; PE-conjugated CD44 (IM7); eFluor450-conjugated CD25 (PC61.5); and APC-eFluor780-conjugated c-KIT. Earliest thymic progenitors were defined as CD4$^-$CD8$^-$ CD11b$^-$B220$^-$TER119$^-$CD3e$^-$GR-1$^-$CD44$^+$CD25$^-$c-KIT$^+$ thymocytes.

Stromal cell analysis and sorting. To extract bone marrow stromal cells, intact bones were crushed in PBS using a mortar and pestle. Crushed fractions in PBS were collected and stored on ice. The bone chips underwent enzymatic digestion by collagenase type II (3 mg ml^{-1}, Worthington Biochemical) and dispase (4 mg ml^{-1}, Roche) at 37 °C for 45 min at 37 °C in a shaking water bath. Both crushed and digested fractions were pooled. Following RBC lysis, endothelial cells were stained with APC-conjugated CD45 (30-F11), FITC-conjugated lineage cocktail (CD3e, GR-1, B220 and TER119) and biotinylated anti-mouse CD31 (PECAM-1), followed by streptavidin PE. Dead cells were excluded using 7-AAD (EMD Biosciences). Perivascular mesenchymal progenitor cells were stained with APC-eFluor780-conjugated CD45 and APC-eFluor780-conjugated lineage cocktail, APC-conjugated SCA1, biotinylated CD140a (PDGRFα), and streptavidin Brilliant Violet 421 (Biolegend). FACS analyses were performed using FACScan (BD Biosciences), LSRII (BD Biosciences) or Gallios (Beckman Coulter) flow cytometers. Cell sorting was done on Reflection (iCyt) or Aria (BD Biosciences) flow cytometers. RNA of sorted cells was extracted using NucleoSpin RNA XS kit (Macherey-Nagel) per the manufacture's recommendations.

Quantitative reverse-transcription PCR. For total bone marrow RNA, femurs were flushed with 1 ml of Trizol (Invitrogen). RNA was prepared according to the manufacturer's specification. One-step quantitative reverse-transcription PCR was performed using the TaqMan Universal PCR Master Mix (Applied Biosystems) using no template and no reverse-transcription controls. Data was collected on a 7300 Real-Time PCR System (Applied Biosystems). Primers were as follows: CXCL12 forward, 5′-GAGCCAACGTCAAGCATCTG-3′; CXCL12 reverse, 5′-CGGGTCAATGCACACTTGTC-3′; CXCL12 dT-FAM/TAMRA probe, 5′-TCCAAACTGTGCCCTTCAGATTGTTGC-3′; β-actin forward, 5′-ACCAACTGGGACGATATGGAGAAGA-3′; β-actin primer; and β-actin dT-VIC/TAMRA probe, 5′-AGCCATGTACGTAGCCATCCAGGCTG-3′.

Immunofluorescence. Mice were lethally sedated and perfused with PBS followed by 4% paraformaldehyde. Hindlimbs were removed and post-fixed in 4% paraformaldehyde overnight at 4 °C. Bones were washed in water, decalcified in 14% EDTA, pH 7.4, and cryoprotected in 30% sucrose in PBS. Bones were then snap frozen in OCT using liquid-nitrogen-cooled 2-methylbutane, and blocks were sectioned at 7 μm. For lineage mapping, sections were washed in PBS and mounted with Prolong Gold Antifade Reagent with DAPI (Invitrogen). Slides were imaged using an ApoTome fluorescent microscope (Zeiss).

CFU-C assay. We plated 25,000 bone marrow cells or 50,000 spleen cells in 2.75 ml methylcellulose media (MethoCult 3434; StemCell Technologies). Whole peripheral blood (20 μl) was RBC-lysed and plated in methylcellulose. Duplicate cultures were incubated at 37 °C for 7 days, after which colonies containing at least 100 cells were counted in a blind fashion.

CFU-F assay. PDGFRα$^+$SCA1$^+$ cells were sorted by flow cytometry and directly plated on tissue culture plates containing alpha MEM and 10% fetal bovine serum (Atlanta Biologicals). Media exchanges were performed every 3–4 days for a total of 14 days, after which colonies containing at least 50 cells were counted. On day 14, cells were harvested from the cultures, replated in osteogenic or adipogenic media, and cultured for an additional 14 days. Osteogenic media: alpha MEM with 10% fetal bovine serum, 50 μM ascorbic acid (Sigma) and 10 μM β-glycerophosphate (Sigma). Adipogenic media: alpha MEM with 10% fetal bovine serum, 5 μg ml^{-1} insulin, 100 μM indomethacin and 100 nM dexamethasone. Osteoblast differentiation was assessed using the Leukocyte Alkaline Phosphatase Kit (Sigma), per the manufacture's recommendations. Adipocyte differentiation was assessed by staining with Oil Red O, as reported previously[32].

RNA expression profiling. PDGFRα$^+$SCA1$^+$ cells or CAR cells, pooled from 2–6 mice, were sorted directly into lysis buffer and RNA was prepared using the RNA XS column kit (Macherey-Nagel, Pennsylvania) according to the manufacturer's directions. RNA was amplified using the NuGen Ovation system (NuGen, California) and hybridized to the Affymetrix MoGene 1.0 ST array. Data normalization was performed using the robust multichip average algorithm.

Statistics. Significance was determined using PRISM software (GraphPad). Unless otherwise stated, the statistical significance of differences was calculated using one- or two-way analysis of variance. *P* values indicate the result of Bonferroni post-testing relative to *Cxcl12*$^{fl/-}$ control mice unless other comparisons are explicitly shown. *P* values less than 0.05 were considered significant. All data are presented as mean ± s.e.m.

30. Lakso, M. *et al.* Efficient in vivo manipulation of mouse genomic sequences at the zygote stage. *Proc. Natl Acad. Sci. USA* **93,** 5860–5865 (1996).
31. Semerad, C. L., Liu, F., Gregory, A. D., Stumpf, K. & Link, D. C. G-CSF is an essential regulator of neutrophil trafficking from the bone marrow to the blood. *Immunity* **17,** 413–423 (2002).
32. Churukian, C. Lillie's oil red O method for neutral lipids. *J. Histotechnol.* **22,** 309–311 (1999).

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature11885

Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches

Lei Ding[1] & Sean J. Morrison[1]

Although haematopoietic stem cells (HSCs) are commonly assumed to reside within a specialized microenvironment, or niche[1], most published experimental manipulations of the HSC niche have affected the function of diverse restricted progenitors. This raises the fundamental question of whether HSCs[1] and restricted progenitors[2,3] reside within distinct, specialized niches or whether they share a common niche. Here we assess the physiological sources of the chemokine CXCL12 for HSC and restricted progenitor maintenance. *Cxcl12*^{DsRed} knock-in mice (DsRed-Express2 recombined into the *Cxcl12* locus) showed that *Cxcl12* was primarily expressed by perivascular stromal cells and, at lower levels, by endothelial cells, osteoblasts and some haematopoietic cells. Conditional deletion of *Cxcl12* from haematopoietic cells or nestin–*cre*-expressing cells had little or no effect on HSCs or restricted progenitors. Deletion of *Cxcl12* from endothelial cells depleted HSCs but not myeloerythroid or lymphoid progenitors. Deletion of *Cxcl12* from perivascular stromal cells depleted HSCs and certain restricted progenitors and mobilized these cells into circulation. Deletion of *Cxcl12* from osteoblasts depleted certain early lymphoid progenitors but not HSCs or myeloerythroid progenitors, and did not mobilize these cells into circulation. Different stem and progenitor cells thus reside in distinct cellular niches in bone marrow: HSCs occupy a perivascular niche and early lymphoid progenitors occupy an endosteal niche.

Using signalling lymphocyte attractant molecule (SLAM) family markers that isolate quiescent HSCs[4–8], we found that most HSCs localize adjacent to sinusoidal blood vessels in the bone marrow[4,9,10]. Using independent approaches, others have obtained similar results[11–13]. We therefore proposed that there is a perivascular niche for HSC maintenance[4]. Consistent with this, stem cell factor (SCF, also known as KITL) is primarily or exclusively expressed in the bone marrow by endothelial cells and perivascular stromal cells[10]. Conditional deletion of *Scf* from endothelial cells or leptin receptor (*Lepr*)-expressing perivascular stromal cells depleted HSCs[10]. Combined deletion of *Scf* from both endothelial cells and perivascular stromal cells caused severe HSC depletion and anaemia. In contrast, conditional deletion of *Scf* from osteoblasts or haematopoietic cells did not affect HSC frequency or function. This proves that there is a perivascular niche for HSC maintenance and raises the question of whether other haematopoietic progenitors reside in distinct niches.

CXCL12 is a chemokine required for HSC maintenance and retention in the bone marrow[11,14–17]. Global conditional deletion of *Cxcl12*, or the gene that encodes its receptor, *Cxcr4*, depletes HSCs from adult bone marrow[11,14]. CXCL12 also promotes the proliferation and maintenance of B-lineage progenitors[15,18] and common lymphoid progenitors[19] (CLPs). CXCL12 is expressed by perivascular stromal cells, endothelial cells and osteoblasts[11,20,21]. Some have proposed that the physiological source of CXCL12 for HSC maintenance is osteoblasts[14], whereas others have proposed that the source is perivascular stromal cells[11]; however, *Cxcl12* has not yet been conditionally deleted from any candidate niche cell. Thus, the cellular sources of CXCL12 for the maintenance of HSCs and lymphoid progenitors remain uncertain.

To examine the *Cxcl12* expression pattern systematically, we generated *Cxcl12*^{DsRed} knock-in mice by recombining DsRed-Express2 (*DsRed*) into the endogenous *Cxcl12* locus (Supplementary Fig. 1a–c). *Cxcl12* was primarily expressed by cells surrounding sinusoids throughout the bone marrow, irrespective of proximity to the endosteum (Fig. 1a–c and Supplementary Fig. 1d). *Cxcl12*–DsRed expression overlapped with endothelial marker staining, suggesting that endothelial cells were one source of CXCL12 (Fig. 1a–c); however, perivascular stromal cells also seemed to produce CXCL12 (Fig. 1a–c).

The perivascular *Cxcl12* expression pattern was very similar to the *Scf* expression pattern[10]. In *Scf*$^{gfp/+}$;*Cxcl12*$^{DsRed/+}$ mice, we found a strong overlap in *Cxcl12*–DsRed and *Scf*–GFP expression by perivascular cells throughout the bone marrow (Fig. 1d–f). From flow cytometry, virtually all *Scf*–GFP^{+} stromal cells were positive for *Cxcl12*–DsRed (Fig. 1j) and virtually all *Cxcl12*–DsRed^{+} stromal cells were positive for *Scf*–GFP (Fig. 1k). Perivascular HSC niche cells therefore produce both SCF and CXCL12.

To quantitate *Cxcl12* expression in perivascular stromal cells, we sorted CD45/TER119^{-}PDGFRα^{+} mesenchymal stem/stromal cells from enzymatically dissociated bone marrow. The *Lepr*–Cre-expressing perivascular stromal cells that contribute to the HSC niche by secreting SCF are uniformly positive for PDGFRα (ref. 10), and virtually all CD45/TER119^{-}PDGFRα^{+} bone marrow cells express *Scf*–GFP (Supplementary Fig. 1e, f). Approximately 90% of CD45/TER119^{-}PDGFRα^{+} perivascular stromal cells expressed *Cxcl12*–DsRed (Fig. 1l). From flow cytometry, approximately 70% of VE-cadherin^{+} endothelial cells and 0.5% of CD45/TER119^{+} haematopoietic cells in the bone marrow also expressed *Cxcl12*–DsRed (Fig. 1m, n and Supplementary Fig. 1g).

In contrast to *Scf*, which is not detectably expressed by osteoblasts[10], we did observe very low levels of *Cxcl12*–DsRed expression by bone-lining cells (Fig. 1g–i). Almost 70% of *Col2.3*–GFP^{+} osteoblasts isolated from enzymatically digested bone from *Col2.3–GFP*;*Cxcl12*$^{DsRed/+}$ mice expressed low levels of *Cxcl12*–DsRed (Fig. 1o).

By quantitative reverse-transcription polymerase chain reaction (qRT–PCR), EYFP^{+} perivascular stromal cells from *Lepr–cre*;*loxP–EYFP* mice expressed *Cxcl12* at ~15,000-fold the level observed in unfractionated bone marrow (Fig. 1p). VE-cadherin^{+} endothelial cells, *Col2.3*–GFP^{+} osteoblasts, and *Cxcl12*–DsRed^{+} haematopoietic cells expressed significantly lower levels of *Cxcl12* at ~120-fold, ~13-fold and ~3-fold the levels observed in bone marrow cells (Fig. 1p).

We generated a floxed allele of *Cxcl12* (*Cxcl12*^{fl}) that led to a frameshift mutation on recombination (Supplementary Fig. 2a–c). The floxed allele itself did not seem to have any phenotype, because unrecombined *Cxcl12*$^{fl/fl}$ mice were born and matured into adulthood in normal numbers with normal HSC frequency and haematopoiesis (Supplementary Fig. 2d–f). We recombined *Cxcl12*^{fl} in the germ line with *CMV–cre* mice to generate a *Cxcl12*$^{-}$ predicted null allele. *Cxcl12*$^{+/-}$ mice were born in expected numbers (Supplementary Fig. 2g) with normal cellularity, B-cell frequency and HSC frequency in the bone marrow and spleen (Supplementary Fig. 2h–j). In contrast, *Cxcl12*$^{-/-}$ progeny were not born alive (Supplementary Fig. 2g)

[1]Howard Hughes Medical Institute, Children's Research Institute, Department of Pediatrics, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390, USA.

©2013 Macmillan Publishers Limited. All rights reserved

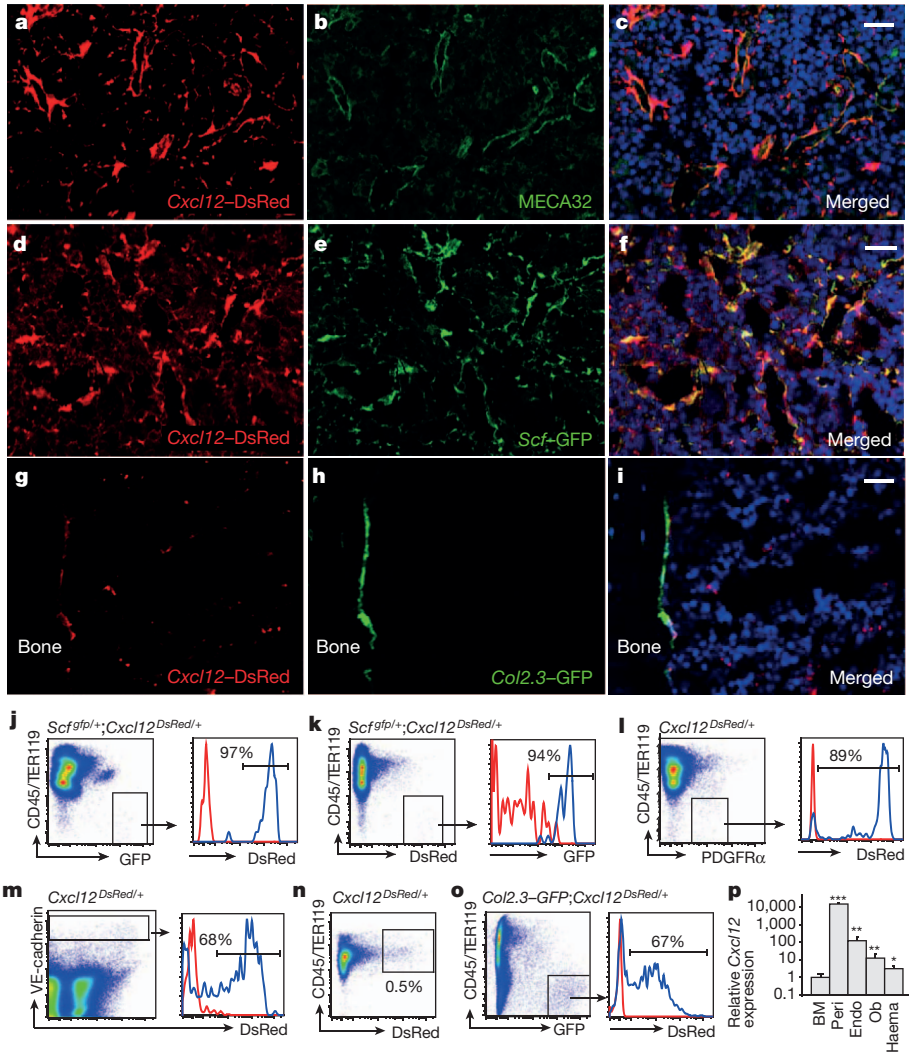


Figure 1 | Endothelial cells and perivascular stromal cells are the major sources of *Cxcl12* in bone marrow. **a**–**c**, In *Cxcl12*$^{DsRed/+}$ mice, DsRed was primarily expressed by perivascular cells throughout the bone marrow. MECA32 also known as PLVAP. **d**–**f**, *Cxcl12*–DsRed and *Scf*–GFP expression strongly overlapped around sinusoids throughout the bone marrow of *Scf*$^{gfp/+}$;*Cxcl12*$^{DsRed/+}$ mice. **g**–**i**, *Col2.3*–GFP^{+} bone-lining osteoblasts expressed *Cxcl12*–DsRed in the bone marrow of *Col2.3–GFP*$^{+}$;*Cxcl12*$^{DsRed/+}$ mice. Nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI, blue) in **c**, **f** and **i**. **j**, CD45/TER119^{-}*Scf*–GFP^{+} perivascular stromal cells expressed *Cxcl12*–DsRed in *Scf*$^{gfp/+}$;*Cxcl12*$^{DsRed/+}$ marrow (blue histogram) but not in *Scf*$^{gfp/+}$ control marrow (red histogram). **k**, CD45/TER119^{-}*Cxcl12*–DsRed^{high} cells expressed *Scf*–GFP in *Scf*$^{gfp/+}$;*Cxcl12*$^{DsRed/+}$ (blue) but not in *Cxcl12*$^{DsRed/+}$ control marrow (red). **l**, CD45/TER119^{-}PDGFRα^{+} perivascular stromal cells expressed *Cxcl12*–DsRed in *Cxcl12*$^{DsRed/+}$ marrow (blue) but not control (+/+) marrow (red). **m**, VE-cadherin^{+} endothelial cells expressed *Cxcl12*–DsRed in *Cxcl12*$^{DsRed/+}$ marrow (blue) but not control (+/+) marrow (red). **n**, 0.5% of CD45/TER119^{+} haematopoietic cells expressed *Cxcl12*–DsRed in *Cxcl12*$^{DsRed/+}$ marrow but not in control marrow (Supplementary Fig. 1g). **o**, CD45/TER119^{-}*Col2.3*–GFP^{+} osteoblasts from enzymatically dissociated bone expressed *Cxcl12*–DsRed in *Col2.3–GFP*;*Cxcl12*$^{DsRed/+}$ mice (blue) but not *Col2.3–GFP* control mice (red). **p**, *Cxcl12* transcript levels from qRT–PCR in different subpopulations of bone marrow cells ($n = 3$–6). BM, whole bone marrow cells; peri, EYFP^{+} perivascular stromal cells from *Lepr–cre*;*loxP–EYFP* mice; endo, VE-cadherin^{+} bone marrow endothelial cells; ob, *Col2.3*–GFP^{+} osteoblasts; haema, CD45/TER119^{+}*Cxcl12*–DsRed^{+} haematopoietic cells. All data represent mean ± s.d. from at least three independent experiments. Two-tailed Student's *t*-tests were used to assess statistical significance (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). Scale bars, 20 μm.

consistent with the known perinatal lethal phenotype of *Cxcl12*-deficient mice[15].

Global deletion of *Cxcl12* by administering tamoxifen to eight-week-old adult *Ubc–creER*;*Cxcl12*$^{fl/fl}$ mice significantly reduced white blood cell counts (Supplementary Fig. 4a), lymphocyte frequencies (Supplementary Fig. 4b), bone marrow cellularity (Supplementary Fig. 4c) and CD150^{+}CD48^{-}LIN^{-}SCA1^{+}c-KIT^{+} HSC[4] frequency (Supplementary Fig. 4d). Bone marrow cells from *Ubc–creER*;*Cxcl12*$^{fl/fl}$ mice also gave significantly lower levels of donor cell reconstitution in all major haematopoietic lineages upon transplantation into irradiated mice (Supplementary Fig. 4e). Consistent with an independently targeted *Cxcl12*^{fl} allele[14], these results demonstrate CXCL12 promotes adult HSC maintenance and lymphopoiesis.

Flow cytometry shows that HSCs do not express *Cxcl12* (Supplementary Fig. 4f). However, because some other haematopoietic cells expressed *Cxcl12*–DsRed (Fig. 1n, p), we conditionally deleted *Cxcl12* from all haematopoietic cells in *Vav1–cre*;*Cxcl12*$^{fl/fl}$ mice. Recombination was highly efficient in *Vav1–cre*;*Cxcl12*$^{fl/fl}$ HSCs (Supplementary Fig. 5a). Adult *Vav1–cre*;*Cxcl12*$^{fl/fl}$ mice had normal cellularity, HSC frequency (Supplementary Fig. 4g, h), lineage composition in the bone marrow and spleen (Supplementary Fig. 5b), and in irradiated mice, reconstituting potential (except for a modest decline in T-cell reconstitution; Supplementary Fig. 4i). Cells from the bone marrow, spleen and blood of *Vav1–cre*;*Cxcl12*$^{fl/fl}$ mice formed normal numbers of haematopoietic colonies in culture (Supplementary Fig. 5c). We were unable to detect HSCs in the blood of *Vav1–cre*;*Cxcl12*$^{fl/fl}$ mice when we competitively transplanted 600,000 mononucleated blood cells into irradiated mice (data not shown).

Adult *Vav1–cre*;*Cxcl12*$^{fl/fl}$ mice also had normal frequencies of CD150^{-}CD48^{-}LSK multipotent progenitors (MPPs; see Supplementary Fig. 3 for flow cytometry gates and references for each cell population), FLT3^{+}LSK lymphoid-primed MPPs (LMPPs), CD34^{+}FcγR^{-}LIN^{-}SCA1^{-}c-KIT^{+} common myeloid progenitors (CMPs), LIN^{-}SCA1^{low}c-KIT^{low}FLT3^{+}IL7Rα^{+} CLPs, CD34^{-}FcγR^{-}LIN^{-}SCA1^{-}c-KIT^{+} megakaryocytic–erythroid progenitors (MEPs), and CD34^{+}FcγR^{+}LIN^{-}SCA1^{-}c-KIT^{+} granulocyte–macrophage progenitors (GMPs) (Supplementary Fig. 4j, k). Committed B-lineage progenitors including B220^{+}sIgM^{-}CD43^{+}CD24^{-} pre-pro B cells, B220^{+}sIgM^{-}CD43^{+}CD24^{+} pro B cells, B220^{+}sIgM^{-}CD43^{-} pre B cells and B220^{+}sIgM^{-} B cells were also similar to controls (Supplementary Fig. 4l). *Cxcl12* expression by haematopoietic cells is therefore not required for the maintenance or retention of HSCs or restricted progenitors in adult bone marrow.

Nestin–*cre*;*Cxcl12*$^{fl/fl}$ mice also had normal bone marrow and spleen cellularity, HSC frequency and haematopoietic lineage composition (Supplementary Fig. 6a, b). Bone marrow cells from nestin–*cre*;*Cxcl12*$^{fl/fl}$ mice were indistinguishable from control cells in their capacity to give long-term multilineage reconstitution of irradiated mice (Supplementary Fig. 6c). The bone marrow of nestin–*cre*;*Cxcl12*$^{fl/fl}$ mice had normal frequencies of MPPs, LMPPs, CMPs, CLPs, MEPs, GMPs and committed B-lineage progenitors (Supplementary Fig. 6d–f). Nestin–*cre*;*Cxcl12*$^{fl/fl}$ mice also had normal

©2013 Macmillan Publishers Limited. All rights reserved

frequencies of colony-forming progenitors in the bone marrow, spleen and blood (Supplementary Fig. 6g–i). We were unable to detect HSCs in the blood of these mice in competitive reconstitution assays (data not shown). Nestin–*cre*-expressing cells are therefore not a physiologically important source of CXCL12 for the maintenance of HSCs or restricted progenitors in the bone marrow. Although HSC niche cells do not express endogenous nestin, nestin–*cre* or nestin–*creER*, they do seem to express the nestin–GFP transgene[10], consistent with previously published results[12].

Given that *Cxcl12* from haematopoietic cells was not required by HSCs or restricted progenitors (Supplementary Fig. 4g–l), *Tie2–cre*;*Cxcl12*$^{fl/fl}$ (*Tie2* also known as *Tek*) mice allowed us to test whether *Cxcl12* from endothelial cells is physiologically important. Adult *Tie2–cre*;*Cxcl12*$^{fl/fl}$ mice had normal bone marrow and spleen cellularity (Fig. 2a), blood cell counts and lineage composition in the bone marrow and spleen (Supplementary Fig. 7a, b). However, the frequency of CD150^{+}CD48^{-}LIN^{-}SCA1^{+}c-KIT^{+} HSCs in the bone marrow of *Tie2–cre*;*Cxcl12*$^{fl/fl}$ mice was significantly lower than in littermate controls (Fig. 2b). Bone marrow cells from *Tie2–cre*;*Cxcl12*$^{fl/fl}$ mice also gave significantly lower levels of donor cell reconstitution in all major haematopoietic lineages on transplantation into irradiated mice (Fig. 2c). Endothelial cells are therefore a physiologically important source of CXCL12 for HSC maintenance. This probably reflects a requirement for CXCL12 produced by postnatal bone marrow endothelial cells because *Cxcl12*-deficient mice do not show HSC depletion during fetal development, only in postnatal bone marrow[17,22]. HSCs from *Tie2–cre*;*Cxcl12*$^{fl/fl}$ mice had normal frequencies of BrdU^{+} cells and Annexin V^{+} cells, suggesting that HSCs may be depleted in these mice through premature differentiation (Supplementary Fig. 7c, d).

Tie2–cre;*Cxcl12*$^{fl/fl}$ mice had normal frequencies of MPPs, LMPPs, CMPs, CLPs, MEPs, GMPs and B-lineage progenitors in the bone marrow (Fig. 2d–f) and colony-forming progenitors in the bone

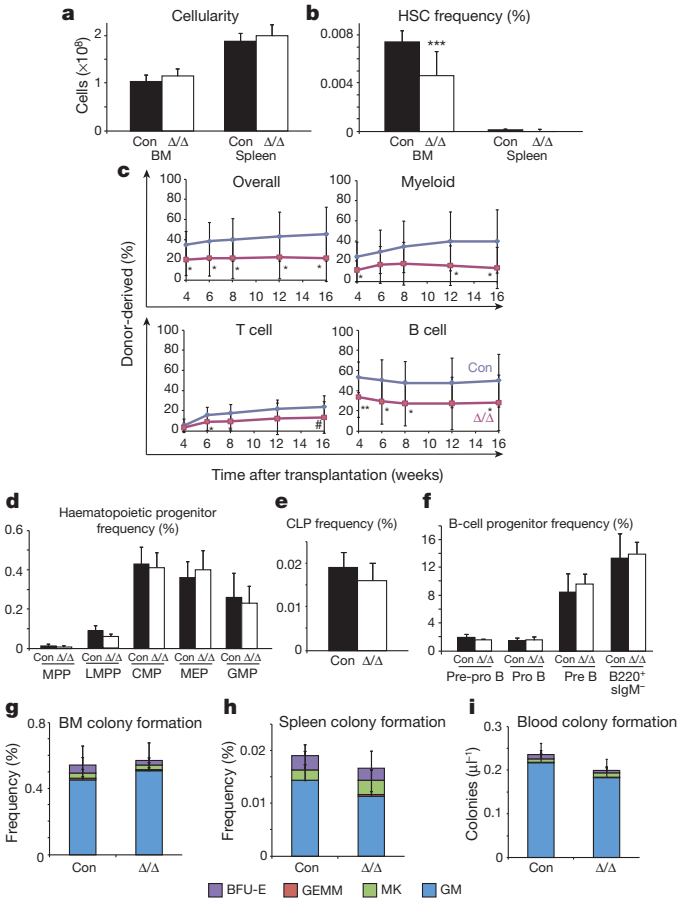


Figure 2 | CXCL12 produced by endothelial cells promotes HSC maintenance. **a**, **b**, Bone marrow and spleen cellularity ($n = 8, 9$) (**a**) and HSC frequency in *Tie2–cre*;*Cxcl12*$^{fl/fl}$ mice versus littermate controls ($n = 10, 11$) (**b**). **c**, 3×10^5 donor bone marrow cells from *Tie2–cre*;*Cxcl12*$^{fl/fl}$ mice gave significantly lower levels of donor myeloid cell, B-cell and T-cell reconstitution in irradiated mice (three experiments with a total of 12–14 recipient mice per genotype) (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, # = 0.057). **d**–**f**, *Tie2–cre*;*Cxcl12*$^{fl/fl}$ mice had normal frequencies of MPPs, LMPPs, CMPs, MEPs and GMPs (**d**), CLPs (**e**) and committed B-lineage progenitors in bone marrow (**f**) ($n = 3, 4$). **g**–**i**, *Tie2–cre*;*Cxcl12*$^{fl/fl}$ mice had normal frequencies of myeloerythroid-colony-forming progenitors in bone marrow (**g**), spleen (**h**) and blood (**i**) ($n = 3$–6). Δ, recombined *Cxcl12*^{fl} allele; con, negative control mice with +/+ or *fl*/+ or *fl*/*fl* *Cxcl12* genotypes (without *cre*). Data are mean ± s.d. BFU-E, erythroid burst-forming unit; MK, megakaryocyte; GEMM, granulocyte–erythrocyte–monocyte–megakaryocyte; GM, granulocyte–macrophage.

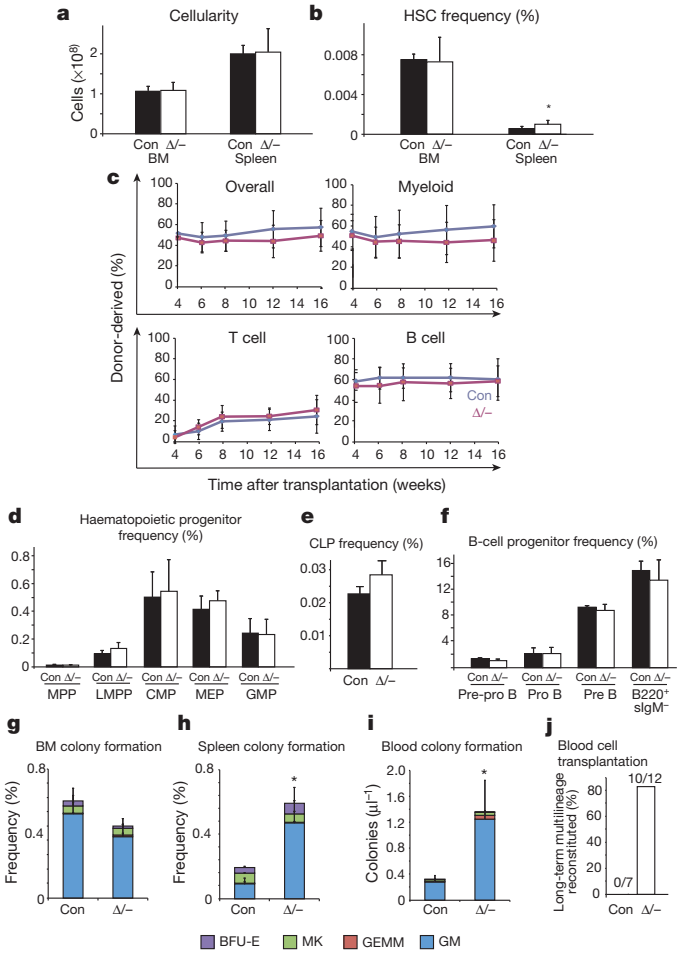


Figure 3 | CXCL12 produced by *Lepr*-expressing perivascular stromal cells retains HSCs and colony-forming progenitors in the bone marrow. **a**, **b**, Cellularity (**a**) ($n = 6$) and HSC frequency (**b**) ($n = 6, 7$) in the bone marrow and spleen of *Lepr–cre*;*Cxcl12*$^{fl/-}$ mice and littermate controls. *$P < 0.05$. **c**, 3×10^5 bone marrow cells from *Lepr–cre*;*Cxcl12*$^{fl/-}$ mice gave normal levels of donor myeloid cell, B-cell and T-cell reconstitution in irradiated mice (three experiments with a total of 15 recipient mice per genotype). **d**–**f**, *Lepr–cre*;*Cxcl12*$^{fl/-}$ mice had normal frequencies of MPPs, LMPPs, CMPs, MEPs and GMPs (**d**), CLPs (**e**) and committed B-lineage progenitors in bone marrow (**f**) ($n = 3$). **g**–**i**, *Lepr–cre*;*Cxcl12*$^{fl/-}$ mice had normal frequencies of myeloerythroid-colony-forming progenitors in bone marrow (**g**) but significantly increased frequencies in spleen (**h**) and blood (**i**) ($n = 3$–5). **j**, 6×10^5 mononucleated blood cells from *Lepr–cre*;*Cxcl12*$^{fl/-}$ mice gave long-term multilineage reconstitution in irradiated mice (10 of 12), whereas blood cells from littermate controls did not (0 of 7). Δ, recombined *Cxcl12*^{fl} allele; −, germline-deleted *Cxcl12* allele or *Cxcl12*^{DsRed} allele; con, control mice. Data represent mean ± s.d.

©2013 Macmillan Publishers Limited. All rights reserved

marrow, spleen and blood (Fig. 2g–i). We were unable to detect HSCs in the blood of these mice in competitive reconstitution assays (data not shown). CXCL12 produced by endothelial cells is therefore not required for the retention of HSCs or most restricted progenitors in bone marrow.

We deleted *Cxcl12* from perivascular stromal cells in the bone marrow using *Lepr–cre* (ref. 10). *Lepr–cre*;*Cxcl12*$^{fl/-}$ mice had normal bone marrow and spleen cellularity (Fig. 3a), normal blood cell counts and normal lineage composition in the bone marrow and spleen (Supplementary Fig. 7e, f). CD150$^+$CD48$^-$LIN$^-$SCA1$^+$c-KIT$^+$ HSC frequency was normal in the bone marrow of *Lepr–cre*;*Cxcl12*$^{fl/-}$ mice (Fig. 3b), and the reconstituting capacity of bone marrow cells in irradiated mice was normal (Fig. 3c). The bone marrow of *Lepr–cre*;*Cxcl12*$^{fl/-}$ mice also had normal frequencies of MPPs, LMPPs, CMPs, CLPs, MEPs, GMPs and B-lineage progenitors (Fig. 3d–f). However, the frequencies of HSCs and colony-forming progenitors were significantly higher in the blood and spleen of *Lepr–cre*;*Cxcl12*$^{fl/-}$ mice than in littermate controls (Fig. 3b, g–j). This demonstrates that CXCL12 expression by *Lepr*$^+$ perivascular cells is required to retain HSCs and colony-forming progenitors in the bone marrow.

Col2.3–Cre recombines genes in fetal and adult osteoblasts[10,23]. *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ mice had normal bone marrow and spleen cellularity, normal CD150$^+$CD48$^-$LIN$^-$SCA1$^+$c-KIT$^+$ HSC frequency in the bone marrow and spleen (Fig. 4a, b), normal blood cell counts and normal lineage composition in the bone marrow and spleen (Supplementary Fig. 8a, b). *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ bone marrow cells gave normal levels of donor myeloid reconstitution but significantly lower levels of T- and B-cell reconstitution in irradiated mice, relative to control cells (Fig. 4c). In contrast, purified HSCs from *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ mice gave levels of reconstitution similar to those of control HSCs in all lineages (Fig. 4d and Supplementary Fig. 8c). This suggests that the reduction in T- and B-cell reconstitution from the bone marrow of *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ mice reflected the depletion of early lymphoid progenitors.

The bone marrow of *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ mice had normal frequencies of MPPs, CMPs, MEPs and GMPs (Fig. 4e) but significantly fewer CLPs and IL7Rα$^+$ LMPPs—cells that are thought to be specified for lymphoid differentiation[24,25] (Fig. 4f, g). The thymus in *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ mice had a normal frequency of early thymic progenitors (B220$^-$GR1$^-$TER119$^-$CD4$^-$CD8$^-$CD44$^+$c-KIT$^+$CD25$^-$ ETPs; Supplementary Fig. 8d, e) and the bone marrow had a normal frequency of committed B-lineage progenitors (Fig. 4h). This suggests that homeostatic mechanisms can restore normal frequencies of downstream lymphoid progenitors despite depletion of primitive lymphoid progenitors. CLPs from *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ mice had normal frequencies of BrdU$^+$ and Annexin V$^+$ cells, suggesting that these cells might be depleted by premature differentiation (Supplementary Fig. 8f, g).

Col2.3–cre;*Cxcl12*$^{fl/fl}$ mice had normal frequencies of myeloerythroid-colony-forming progenitors in the bone marrow, spleen and blood (Fig. 4i), and we were unable to detect HSCs in the blood of these mice in competitive reconstitution assays (data not shown). CXCL12 from osteoblasts is thus required for the maintenance of certain early lymphoid progenitors but not for the maintenance or bone marrow retention of HSCs or myeloerythroid progenitors.

In stained bone marrow sections, we found 30% of LIN$^-$IL7Rα$^+$ cells at the endosteum, adjacent to bone-lining cells (Fig. 4j and Supplementary Fig. 9). Only 5% of all bone marrow cells were adjacent

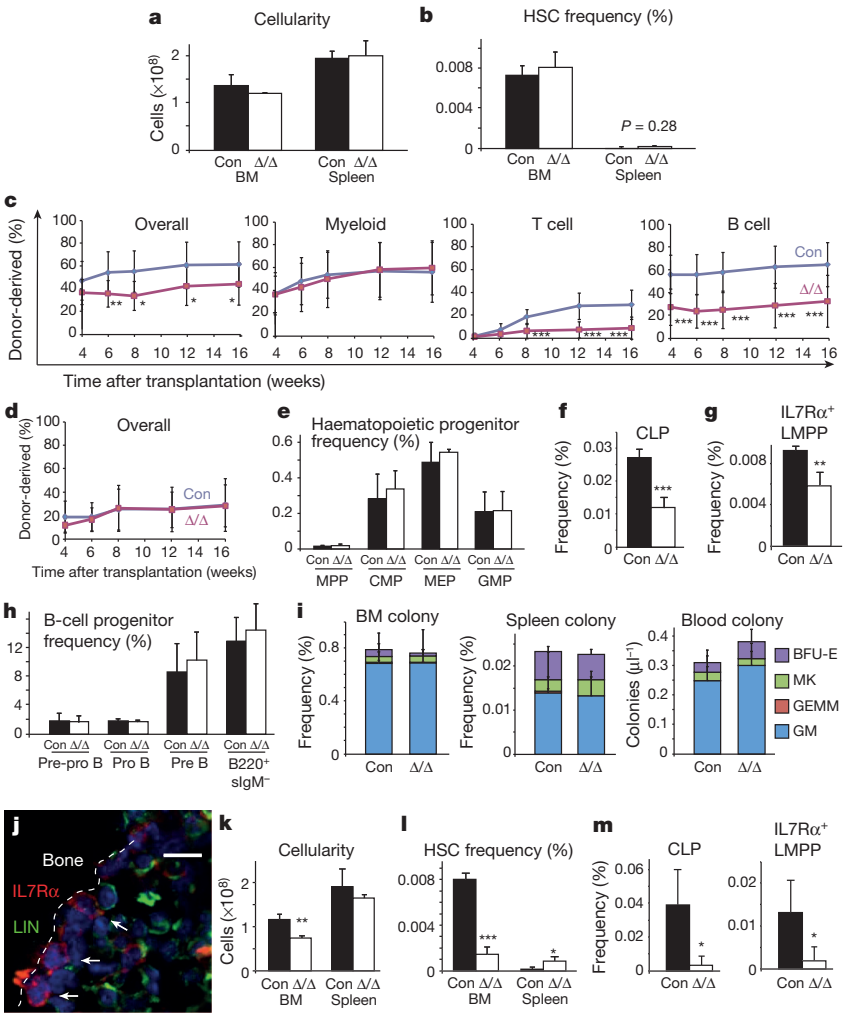


Figure 4 | CXCL12 produced by osteoblasts promotes the maintenance of early lymphoid progenitors but not HSCs. **a**, **b**, Cellularity (**a**) ($n = 4$) and HSC frequency (**b**) ($n = 4$) in the bone marrow and spleen of *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ mice and littermate controls. **c**, 3×10^5 bone marrow cells from *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ mice gave significantly lower levels of donor cell reconstitution in the T- and B-cell lineages but not in the myeloid lineage relative to control bone marrow cells (three experiments with a total of 13 or 14 recipient mice per genotype) (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). **d**, Twenty CD150$^+$CD48$^-$LIN$^-$SCA1$^+$c-KIT$^+$ HSCs from *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ mice gave normal donor cell reconstitution (three experiments with a total of 14 or 15 recipient mice per genotype), including normal levels of myeloid cells, B cells and T cells (Supplementary Fig. 8c). **e**–**h**, *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ bone marrow had normal frequencies of MPPs, CMPs, MEPs and GMPs (**e**) and committed B-lineage progenitors (**h**), but significantly reduced frequencies of CLPs (**f**) and IL7Rα$^+$ LMPPs (**g**) ($n = 3$–5). **i**, *Col2.3–cre*;*Cxcl12*$^{fl/fl}$ mice had normal frequencies of myeloerythroid-colony-forming progenitors in the bone marrow, spleen and blood ($n = 3$–6). **j**, Some LIN$^-$IL7Rα$^+$ early lymphoid progenitors were adjacent to the endosteum. **k**–**m**, *Prx1–cre*;*Cxcl12*$^{fl/fl}$ mice showed significant reductions in bone marrow cellularity (**k**) ($n = 3, 4$) and the frequencies of HSCs (**l**), CLPs and IL7Rα$^+$ LMPPs (**m**) ($n = 4, 5$). Δ, recombined *Cxcl12*^{fl} allele; con, control mice. Data represent mean ± s.d.

©2013 Macmillan Publishers Limited. All rights reserved

to the endosteum in the sections we examined, so these early lymphoid progenitors were six times more likely than the average bone marrow cell to be adjacent to endosteum. Both functional genetic data and localization data therefore suggest that there is an endosteal niche for a subset of early lymphoid progenitors.

To test whether both HSCs and early lymphoid progenitors would be depleted by deleting *Cxcl12* in perivascular stromal cells and in osteoblasts, we used *Prx1–cre* (ref. 26; *Prx1* also known as *Prrx1*). *Prx1*–Cre deletes broadly in perivascular stromal cells as well as in osteoblasts[27]. Consistent with this, we found that *Prx1*–Cre recombined a conditional reporter in osteoblasts as well as in $95 \pm 4\%$ CD45/TER119$^-$PDGFRα^+ perivascular stromal cells (Supplementary Fig. 10a–c), but not in bone marrow endothelial cells (Supplementary Fig. 10d). Thus, *Prx1*–Cre recombined more broadly in perivascular stromal cells than did *Lepr*–Cre, which recombined a conditional reporter in $70 \pm 9\%$ of CD45/TER119$^-$PDGFRα^+ cells (Supplementary Fig. 1f). Consistent with our results with other Cre alleles, *Prx1–cre*;*Cxcl12*$^{fl/fl}$ mice showed significant reductions in bone marrow cellularity (Fig. 4k) and in the frequencies of HSCs (Fig. 4l), CLPs and IL7Rα^+ LMPPs (Fig. 4m). The bone marrow of *Prx1–cre*;*Cxcl12*$^{fl/fl}$ mice also had fewer committed B-lineage progenitors, but CMPs, MEPs and GMPs were normal (Supplementary Fig. 10e, f).

Our data indicate that HSCs depend on a perivascular niche created by endothelial cells and *Lepr–cre*- or *Prx1–cre*-expressing perivascular stromal cells, whereas some early lymphoid progenitors depend on an endosteal niche created by osteoblasts, and committed B-lineage progenitors depend on a distinct perivascular niche created by *Prx1–cre*-expressing stromal cells but not endothelial cells. Similar conclusions have been reached by another group[27]. Osteoblasts promote the proliferation and differentiation of lymphoid progenitors in culture, and conditional ablation of osteoblasts *in vivo* acutely depletes lymphoid progenitors but not HSCs[28,29]. Conditional deletion of stimulatory G-protein α-subunit from osteoblasts depletes B-lineage progenitors but not other haematopoietic progenitors[30]. The data are thus consistent with a number of previous studies, including our recent work[10], in suggesting that osteoblasts create a niche for certain early lymphoid progenitors but not for HSCs.

METHODS SUMMARY

Targeting vectors for making *Cxcl12*^{DsRed} and *Cxcl12*^{fl} mice were generated by recombineering and electroporated into C57BL-derived Bruce4 ES cells. Chimaeric mice were bred with C57BL/6-Tyr$^{c\text{-}2J}$ mice to obtain germline transmission. The *Frt*-flanked *Neo* cassette was deleted by mating with *Flpe* mice (Supplementary Materials). *cre* alleles used in this study were almost all obtained from The Jackson Laboratory, Maine (Supplementary Materials). *Scf*$^{gfp/+}$ mice were described previously[10]. All mice were backcrossed and maintained on a C57BL/Ka background.

Full Methods and any associated references are available in the online version of the paper.

Received 7 August 2012; accepted 3 January 2013.
Published online 24 February 2013.

1. Morrison, S. J. & Spradling, A. C. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* **132,** 598–611 (2008).
2. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404,** 193–197 (2000).
3. Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91,** 661–672 (1997).
4. Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C. & Morrison, S. J. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121,** 1109–1121 (2005).
5. Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135,** 1118–1129 (2008).
6. Morita, Y., Ema, H. & Nakauchi, H. Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. *J. Exp. Med.* **207,** 1173–1182 (2010).
7. Kiel, M. J. *et al.* Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain BrdU. *Nature* **449,** 238–242 (2007).
8. Foudi, A. *et al.* Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nature Biotechnol.* **27,** 84–90 (2009).
9. Kiel, M. J., Radice, G. L. & Morrison, S. J. Lack of evidence that hematopoietic stem cells depend on N-cadherin-mediated adhesion to osteoblasts for their maintenance. *Cell Stem Cell* **1,** 204–217 (2007).
10. Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481,** 457–462 (2012).
11. Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25,** 977–988 (2006).
12. Mendéz-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466,** 829–834 (2010).
13. Lo Celso, C. *et al.* Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche. *Nature* **457,** 92–96 (2009).
14. Tzeng, Y. S. *et al.* Loss of Cxcl12/Sdf-1 in adult mice decreases the quiescent state of hematopoietic stem/progenitor cells and alters the pattern of hematopoietic regeneration after myelosuppression. *Blood* **117,** 429–439 (2011).
15. Nagasawa, T. *et al.* Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature* **382,** 635–638 (1996).
16. Petit, I. *et al.* G-CSF induces stem cell mobilization by decreasing bone marrow SDF-1 and up-regulating CXCR4. *Nature Immunol.* **3,** 687–694 (2002).
17. Ara, T. *et al.* Long-term hematopoietic stem cells require stromal cell-derived factor-1 for colonizing bone marrow during ontogeny. *Immunity* **19,** 257–267 (2003).
18. Nagasawa, T., Kikutani, H. & Kishimoto, T. Molecular cloning and structure of a pre-B-cell growth-stimulating factor. *Proc. Natl Acad. Sci. USA* **91,** 2305–2309 (1994).
19. Nie, Y., Han, Y. C. & Zou, Y. R. CXCR4 is required for the quiescence of primitive hematopoietic cells. *J. Exp. Med.* **205,** 777–783 (2008).
20. Ponomaryov, T. *et al.* Induction of the chemokine stromal-derived factor-1 following DNA damage improves human stem cell function. *J. Clin. Invest.* **106,** 1331–1339 (2000).
21. Dar, A. *et al.* Chemokine receptor CXCR4-dependent internalization and resecretion of functional chemokine SDF-1 by bone marrow endothelial and stromal cells. *Nature Immunol.* **6,** 1038–1046 (2005).
22. Kawabata, K. *et al.* A cell-autonomous requirement for CXCR4 in long-term lymphoid and myeloid reconstitution. *Proc. Natl Acad. Sci. USA* **96,** 5663–5667 (1999).
23. Liu, F. *et al.* Expression and activity of osteoblast-targeted Cre recombinase transgenes in murine skeletal tissues. *Int. J. Dev. Biol.* **48,** 645–653 (2004).
24. Adolfsson, J. *et al.* Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* **121,** 295–306 (2005).
25. Inlay, M. A. *et al.* Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes Dev.* **23,** 2376–2381 (2009).
26. Logan, M. *et al.* Expression of Cre recombinase in the developing mouse limb bud driven by a Prxl enhancer. *Genesis* **33,** 77–80 (2002).
27. Greenbaum, A. *et al.* CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature* http://dx.doi.org/10.1038/nature11926 (2013).
28. Visnjic, D. *et al.* Hematopoiesis is severely altered in mice with an induced osteoblast deficiency. *Blood* **103,** 3258–3264 (2004).
29. Zhu, J. *et al.* Osteoblasts support B lymphocyte commitment and differentiation from hematopoietic stem cells. *Blood* **109,** 3706–3712 (2007).
30. Wu, J. Y. *et al.* Osteoblastic regulation of B lymphopoiesis is mediated by $G_S\alpha$-dependent signaling pathways. *Proc. Natl Acad. Sci. USA* **105,** 16976–16981 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the Howard Hughes Medical Institute (HHMI) and the US National Heart, Lung and Blood Institute (5R01-HL097760). L.D. was supported by a Helen Hay Whitney Foundation Fellowship and by the HHMI. We thank M. White and D. Adams for flow cytometry, E. Hughes and T. Saunders at the University of Michigan transgenic core for helping to generate *Cxcl12*^{DsRed} and *Cxcl12*^{fl} mice, and S. Grove, R. Coolon, S. Manning, M. Gross and K. Correll for managing the mouse colony.

Author Contributions L.D. performed all of the experiments. L.D. and S.J.M. conceived the project, designed the experiments, interpreted the results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.M. (sean.morrison@utsouthwestern.edu).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Mice. Targeting vectors for making *Cxcl12*^{DsRed} and *Cxcl12*^{fl} mice were generated by recombineering[31]. Linearized targeting vector was electroporated into C57BL-derived Bruce4 ES cells. Correctly targeted clones were identified by Southern blotting. Following expansion, these ES cell clones were injected into C57BL/6-Tyr$^{c\text{-}2J}$ blastocysts. Chimaeric mice were bred with C57BL/6-Tyr$^{c\text{-}2J}$ mice to obtain germline transmission. The *Frt*-flanked *Neo* selection cassette was removed by subsequent mating with *Flpe* mice[32]. The resulting mice were backcrossed for at least three generations onto a C57BL/Ka background. Mice used in this study included *Ubc–creER* (ref. 33), *CMV–cre* (ref. 34), *Vav1–cre* (ref. 35), nestin–*cre* (ref. 36), *Tie2–cre* (ref. 37), *Lepr–cre* (ref. 38), *Prx1–cre* (ref. 26) and *LoxP–EYFP* (ref. 39) (all from The Jackson Laboratory). *Scf*$^{gfp/+}$ mice were described previously[10]. *Col2.3–cre* mice[23] were obtained from Drs F. Liu and B. Kream (University of Connecticut). *Col2.3*–GFP mice[40] were obtained from Dr D. Rowe. Tamoxifen chow (Harlan) containing tamoxifen citrate (Spectrum Chemical) at 40 mg kg^{-1}, with 5% sucrose added, was administrated to mice for 2–4 months to induce recombination by CreER.

All mice were housed in the Unit for Laboratory Animal Medicine at the University of Michigan or in the Animal Resource Center at the University of Texas Southwestern Medical Center. All protocols were approved by the University of Michigan Committee on the Use and Care of Animals and by the University of Texas Southwestern Institutional Animal Care and Use Committee.

PCR genotyping. Primers for genotyping of the *Cxcl12*^{DsRed} allele: OLD322, 5′-GATGCGCGTCAGAGACCCC-3′; OLD355, 5′-GCGCAGAGCTGCGAGCC TTTC-3′; OLD 415, 5′-CTTGAAGCGCATGAAGGGCTTG-3′. Primers for genotyping of the *Cxcl12*^{fl} allele: OLD312, 5′-GAGCCCAGAACTCTGCCACC-3′; OLD313, 5′-TCTTGCAAAGACCATCCCCTC-3′. Primers for genotyping of the *Cxcl12*$^{-}$ allele: OLD312, 5′-GAGCCCAGAACTCTGCCACC-3′; OLD353, 5′-ACTCTGAGTGAGATATTCACATC-3′.

Long-term competitive reconstitution assay. Adult mice were administered a minimum lethal dose of radiation using a Cesium 137 GammaCell40 Irradiator (MDS) or an XRAD 320 X-ray irradiator (Precision X-Ray Inc.) to deliver two doses of 540 rad (1,080 rad in total) at least 2 h apart. C57BL/6-SJL (CD45.1) mice were used as recipient mice in transplantation experiments. Cells were injected into the retro-orbital venous sinus of anaesthetized mice. Blood cells were subjected to Ficoll centrifugation before transplantation to isolate mononucleated cells. Mononucleated cells were counted before transplantation using either a CBC machine or a haemocytometer with Turks solution. In most experiments, recipient mice were maintained on antibiotic water (1.11 g l^{-1} neomycin sulphate and 0.121 g l^{-1} polymixin B) for 14 d after transplantation and then switched to regular water. Recipient mice were bled from 4 to 16 weeks after transplantation to examine the levels of donor-derived myeloid, B and T cells in their blood. Red blood cells were lysed with ammonium chloride potassium buffer before antibody staining. The antibodies used to analyse donor chimaerism in the blood were anti-CD45.1 (A20), anti-CD45.2 (104), anti-GR1 (8C5), anti-Mac1 (M1/70), anti-B220 (6B2) and anti-CD3 (KT31.1).

Flow cytometry. Bone marrow cells were isolated by flushing the long bones or by crushing the long bones with mortal and pestle in Ca^{2+}- and Mg^{2+}-free HBSS (HBSS−) with 2% heat-inactivated bovine serum. Spleen cells were obtained by crushing the spleen between two glass slides. The cells were dissociated to a single cell suspension by gently passing them through a 25-gauge needle and then filtering them through a 70-μm nylon mesh. The antibodies anti-CD150 (TC15-12F12.2), anti-CD48 (HM48-1), anti-CD41 (MWReg30), anti-SCA1 (E13-161.7), anti-c-KIT (2B8) were used to isolate HSCs, along with the following antibodies against lineage markers: anti-TER119, anti-B220 (6B2), anti-GR1 (8C5), anti-CD2 (RM2-5), anti-CD3 (17A2), anti-CD5 (53-7.3) and anti-CD8 (53-6.7). Haematopoietic progenitors were identified by flow cytometry using the antibodies anti-SCA1 (E13-161.7) and anti-c-KIT (2B8), along with the following antibodies against lineage markers: anti-TER119, anti-B220 (6B2), anti-GR1 (8C5), anti-CD2 (RM2-5), anti-CD3 (17A2), anti-CD5 (53-7.3) and anti-CD8 (53-6.7), anti-CD34 (RAM34), anti-CD135 (FLT3) (A2F10), anti-CD16/32 (FcγR) (93), anti-CD127 (IL7Rα) (A7R34), anti-CD24 (M1/69), anti-CD43 (1B11), anti-B220 (6B2), anti-IgM (II/41), anti-CD3 (17A2), anti-GR1 (8C5), anti-Mac1 (M1/70), anti-CD41 (MWReg30), anti-CD71 (C2), anti-TER119, anti-CD44 (IM7) and anti-CD25 (PC61). DAPI was used to exclude dead cells.

For flow cytometric analysis of *Scf*-GFP^{+} or *Lepr–cre;loxP–EYFP*$^{+}$ stromal cells, bone marrow was flushed using HBSS− with 2% bovine serum. Then whole bone marrow was digested with Collagenase IV (200 U ml^{-1}) and DNase I (200 U ml^{-1}) at 37 °C for 15 min. Samples were then stained with antibodies and analysed by flow cytometry. Anti-CD140a (APA5), anti-CD45 (30F-11) and anti-TER119 antibodies were used to isolate perivascular stromal cells. For analysis of bone marrow endothelial cells, mice were intravenously injected with 10 μg per mouse of Alexa Fluor 647 conjugated anti-VE-cadherin antibody[41] (BV13, eBiosciences). Ten minutes later, the long bones were removed and bone marrow was flushed, digested and stained as above. Samples were run on FACSAria or FACSCanto II flow cytometers. Data were analysed by FACSDiva (BD) or FlowJo (Tree Star) software. Osteoblasts were dissociated as previously described[42].

Methylcellulose cultures. Cells were sorted into methylcellulose culture medium (3434, Stemcell Technology) and incubated at 37 °C as previously described[9]. Blood cells were obtained by Ficoll centrifuge according to the manufacture's recommendations before plating (GE Healthcare).

Immunostaining bone sections. Freshly dissected long bones were fixed in a formalin-based fixative at 4 °C for 3 h. Then the bones were embedded in 8% gelatine in PBS. Samples were snap-frozen with liquid nitrogen and stored at −80 °C. Bones were sectioned using CryoJane (Instrumedics). Sections were dried overnight at room temperature (~20 °C) and stored at −80 °C. Sections were rehydrated in PBS for 5 min before immunostaining. Goat serum (5%) in PBS was used for blocking. Primary antibodies were applied to the slides for 1 h at room temperature, and this was followed by secondary antibody incubation for 30 min at room temperature with repeated washes in between. Slides were mounted with anti-fade prolong gold (Invitrogen) and images were acquired on an Olympus IX81 microscope or a Zeiss LSM 780 confocal microscope. Primary antibodies were chicken-anti-GFP (Aves), rat-anti-mouse pan-endothelial cell antigen (MECA32, Biolegend), rat-anti-endoglin (eBioscience), rat-anti-RFP (Allele Biotechnology) and goat-anti-osteopontin (R&D). For staining LIN^{-}IL7Rα^{+} cells in bone marrow sections, mice were intravenously injected with 2 μg Alexa Fluor 647 conjugated monoclonal antibody against IL7Rα (A7R34, Biolegend). After 5 min, the mice were killed and long bones were dissected, fixed, sectioned and stained with antibodies against LIN markers (CD2, CD3, CD5, CD8, TER119, GR1 and B220) as described above. Images were acquired using a Zeiss LSM 780 confocal microscope.

Quantitative reverse-transcription PCR. Cells were sorted directly into TRIzol. Total RNA was extracted according to the manufacture's instructions. Total RNA was subjected to reverse transcription using SuperScript III (Invitrogen). Quantitative real-time PCR was run using SYBR green on a LightCycler 480 (Roche). β-actin was used to normalize the RNA content of samples. Primers used in this study: *Cxcl12*: OLD35, 5′-TGCATCAGTGACGGTAAACCA-3′ and OLD36, 5′-GTTGTTCTTCAGCCGTGCAA-3′; β-actin: OLD27, 5′-GCTCT TTTCCAGCCTTCCTT-3′ and OLD28, 5′-CTTCTGCATCCTGTCAGCAA-3′.

31. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13,** 476–484 (2003).
32. Rodriguez, C. I. *et al.* High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nature Genet.* **25,** 139–140 (2000).
33. Ruzankina, Y. *et al.* Deletion of the developmentally essential gene ATR in adult mice leads to age-related phenotypes and stem cell loss. *Cell Stem Cell* **1,** 113–126 (2007).
34. Schwenk, F., Baron, U. & Rajewsky, K. A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23,** 5080–5081 (1995).
35. de Boer, J. *et al.* Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur. J. Immunol.* **33,** 314–325 (2003).
36. Tronche, F. *et al.* Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. *Nature Genet.* **23,** 99–103 (1999).
37. Koni, P. A. *et al.* Conditional vascular cell adhesion molecule 1 deletion in mice: impaired lymphocyte migration to bone marrow. *J. Exp. Med.* **193,** 741–754 (2001).
38. DeFalco, J. *et al.* Virus-assisted mapping of neural inputs to a feeding center in the hypothalamus. *Science* **291,** 2608–2613 (2001).
39. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1,** 4 (2001).
40. Kalajzic, Z. *et al.* Directing the expression of a green fluorescent protein transgene in differentiated osteoblasts: comparison between rat type I collagen and rat osteocalcin promoters. *Bone* **31,** 654–660 (2002).
41. Butler, J. M. *et al.* Endothelial cells are essential for the self-renewal and repopulation of Notch-dependent hematopoietic stem cells. *Cell Stem Cell* **6,** 251–264 (2010).
42. Semerad, C. L. *et al.* G-CSF potently inhibits osteoblast activity and CXCL12 mRNA expression in the bone marrow. *Blood* **106,** 3020–3027 (2005).

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature11918

PRC1 coordinates timing of sexual differentiation of female primordial germ cells

Shihori Yokobayashi[1]†, Ching-Yeu Liang[1,2], Hubertus Kohler[1], Peter Nestorov[1,2], Zichuan Liu[1], Miguel Vidal[3], Maarten van Lohuizen[4], Tim C. Roloff[1] & Antoine H. F. M. Peters[1,2]

In mammals, sex differentiation of primordial germ cells (PGCs) is determined by extrinsic cues from the environment[1]. In mouse female PGCs, expression of stimulated by retinoic acid gene 8 (*Stra8*) and meiosis are induced in response to retinoic acid provided from the mesonephroi[2–5]. Given the widespread role of retinoic acid signalling during development[6,7], the molecular mechanisms that enable PGCs to express *Stra8* and enter meiosis in a timely manner are unknown[2,8]. Here we identify gene-dosage-dependent roles in PGC development for *Ring1* and *Rnf2*, two central components of the Polycomb repressive complex 1 (PRC1)[9,10]. Both paralogues are essential for PGC development between days 10.5 and 11.5 of gestation. *Rnf2* is subsequently required in female PGCs to maintain high levels of *Oct4* (also known as *Pou5f1*) and *Nanog* expression[11], and to prevent premature induction of meiotic gene expression and entry into meiotic prophase. Chemical inhibition of retinoic acid signalling partially suppresses precocious *Oct4* downregulation and *Stra8* activation in *Rnf2*-deficient female PGCs. Chromatin immunoprecipitation analyses show that *Stra8* is a direct target of PRC1 and PRC2 in PGCs. These data demonstrate the importance of PRC1 gene dosage in PGC development and in coordinating the timing of sex differentiation of female PGCs by antagonizing extrinsic retinoic acid signalling.

In mammalian somatic cells, PRC1 and PRC2 proteins are transcriptional repressors that function in large multiprotein complexes and that modify chromatin by mono-ubiquitinating histone H2A at lysine 119 (H2AK119u1) and trimethylating histone H3 at lysine 27 (H3K27me3), respectively[9,12]. At day 12.5 of embryonic development (E12.5), in PGCs marked by Cdh1 (E-cadherin) staining[13], we observed nuclear localization of PRC1 components Rnf2 (also known as Ring1B), Mel18 (also known as Pcgf2) and Rybp (Fig. 1a and Supplementary Fig. 1) as well as a robust H2AK119u1 signal, suggesting the presence of catalytically active PRC1 complexes (Fig. 1a). To

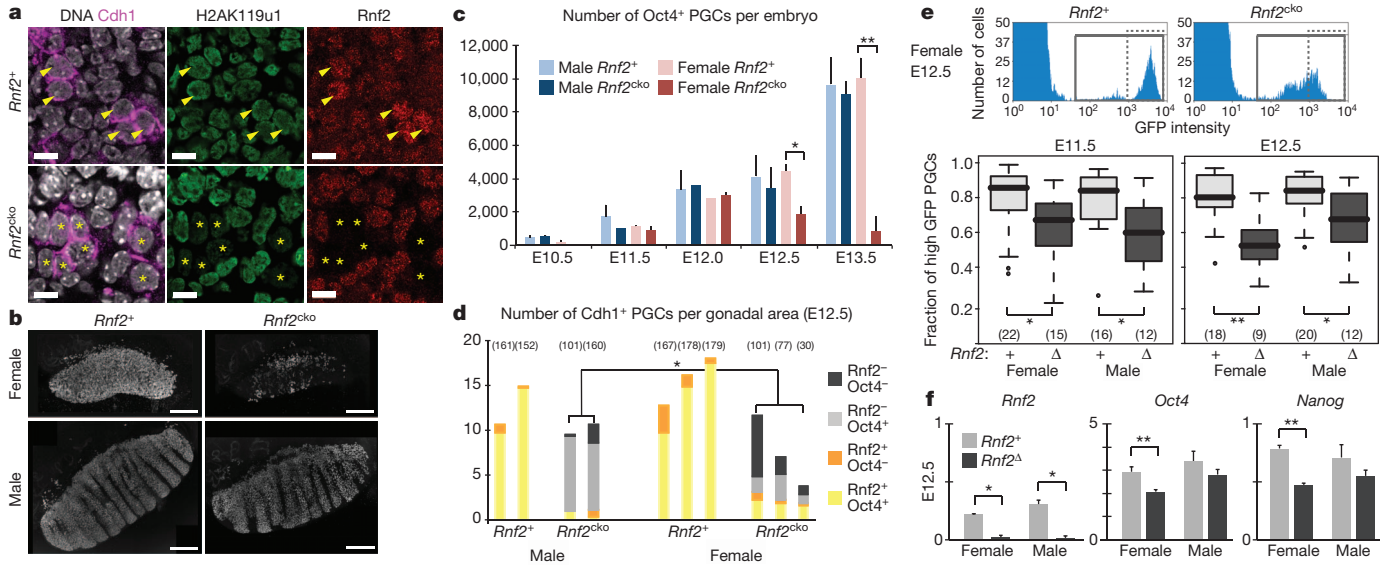


Figure 1 | *Rnf2* regulates PGC development and *Oct4* and *Nanog* expression in *Rnf2*^{cko} female gonads. **a**, Immunofluorescence staining of H2AK119u1, Rnf2 and Cdh1 with DAPI (4′,6-diamidino-2-phenylindole) in *Rnf2*$^{+}$ and *Rnf2*^{cko} gonadal sections from E12.5 female embryos. Arrowheads denote *Rnf2*$^{+}$ PGCs and asterisks denote *Rnf2*$^{\Delta}$ PGCs. Scale bars, 10 μm. **b**, Immunofluorescence staining of Oct4 in E13.5 *Rnf2*$^{+}$ and *Rnf2*^{cko} whole gonads and mesonephroi. Scale bars, 300 μm. **c**, Average number of Oct4-positive cells in whole gonads at E10.5–E13.5. Error bars indicate +s.d. $n = 2$–12. $*P < 0.005$; $**P < 1.0 \times 10^{-5}$ (Student's *t*-test). **d**, Classification of Cdh1-positive PGCs according to Rnf2 and Oct4 protein levels in *Rnf2*$^{+}$ and *Rnf2*^{cko} E12.5 gonads. *y* axis represents the number of PGCs that were normalized to areas analysed (10,000 μm^2). Numbers in brackets denote number of PGCs scored per embryo. $*P < 1.0 \times 10^{-8}$ (chi-squared test). **e**, Representative histograms showing *Oct4*(ΔPE)–GFP signals in PGCs from female *Rnf2*$^{+}$ and *Rnf2*^{cko} E12.5 gonads. Boxplots showing the ratios of PGCs with high GFP intensity (>10^3, enclosed by dashed line in histogram) over all GFP-positive cells (enclosed by solid line) in different embryos. Numbers in brackets denote number of embryos analysed. $*P < 0.05$; $**P < 0.005$ (Student's *t*-test). **f**, Representative qRT–PCR data of *Rnf2*, *Oct4* and *Nanog* in *Rnf2*$^{+}$ and *Rnf2*$^{\Delta}$ PGCs (normalized to *Tbp*). Error bars indicate +s.d. of 2–3 technical replicates. $*P < 0.05$; $**P < 0.01$ (Student's *t*-test).

[1]Friedrich Miescher Institute for Biomedical Research (FMI), Maulbeerstrasse 66, CH-4058 Basel, Switzerland. [2]Faculty of Sciences, University of Basel, CH-4056 Basel, Switzerland. [3]Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas (CSIC), 28040 Madrid, Spain. [4]Division of Molecular Genetics and Centre for Biomedical Genetics, the Netherlands Cancer Institute (NKI), 1066 CX Amsterdam, the Netherlands. †Present address: Department of Reprogramming Science, Center for iPS Cell Research and Application, Kyoto University, 53 Kawahara-cho, Shogoin Yoshida, Sakyo-ku, Kyoto 606-8507, Japan.

©2013 Macmillan Publishers Limited. All rights reserved

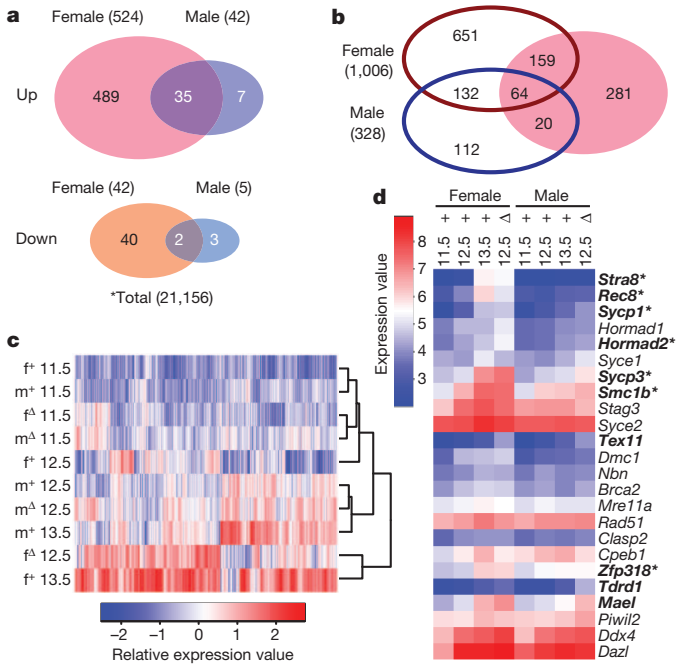


Figure 2 | *Rnf2* deficiency induces extensive transcriptional misregulation in female PGCs. a, Venn diagrams showing the numbers of genes misregulated in male and female $Rnf2^{\Delta}$ PGCs compared to $Rnf2^{+}$ PGCs at E12.5. **b**, Venn diagram showing the numbers of genes upregulated in female (red) and male (blue) control PGCs between E11.5 and E13.5 and in female $Rnf2^{\Delta}$ PGCs compared to $Rnf2^{+}$ PGCs at E12.5 (pink oval). **c**, Relative expression levels of 1138 probe sets upregulated in male (m) and female (f) $Rnf2^{+}$ PGCs between E11.5 and E13.5 (**b**) in various samples indicated. Unsupervised clustering analysis shows clustering of female E12.5 $Rnf2^{\Delta}$ PGCs with female E13.5 $Rnf2^{+}$ PGCs, whereas E12.5 male $Rnf2^{\Delta}$ and $Rnf2^{+}$ PGCs clustered together. **d**, Microarray expression values of early meiosis program genes. Genes significantly upregulated in $Rnf2^{\Delta}$ PGCs of both sexes are indicated in bold; those in female $Rnf2^{\Delta}$ PGCs only in bold with asterisk. $n = 3$ per condition; fold change >1.5; adjusted P value <0.05.

address the function of PRC1 in PGC development (Supplementary Fig. 2), we conditionally deleted *Rnf2* in PGCs from E9.5 onwards by generating mice carrying a floxed and a mutant allele of *Rnf2* ($Rnf2^{fl/\Delta}$) and a Cre recombinase driven by the mouse *Tnap* promoter (*Tnap-cre*)[14,15]. To concomitantly assess possible functional redundancy with the *Rnf2* paralogue *Ring1* (refs 10, 16), we studied mice that were either heterozygous or homozygous deficient for *Ring1* (ref. 17) (referred to as *Rnf2* conditional knockout ($Rnf2^{cko}$) mice and $Ring1^{\Delta}$ $Rnf2^{cko}$ mice, respectively) (Supplementary Figs 3 and 6a). This strategy resulted in ~90% deletion efficiency at E11.5 (Supplementary Fig. 4). At E12.5, Rnf2, Mel18, Rybp and H2AK119u1 were lost in PGCs of $Rnf2^{cko}$ embryos but not of $Rnf2^{fl/\Delta}$ embryos, carrying one functional allele, indicating that complex stability and catalytic activity of PRC1 is regulated by Rnf2 in PGCs at E12.5 (Fig. 1a and Supplementary Figs 1 and 4). By contrast, Ezh2 and H3K27me3 levels were similar in $Rnf2^{cko}$ versus control PGCs, suggesting globally unaltered PRC2 function in $Rnf2^{cko}$ PGCs (Supplementary Fig. 5).

To study the fate of *Rnf2*-deficient PGCs, we analysed expression of the pluripotency and germ cell marker Oct4 in whole gonads of E10.5–E13.5 embryos (Fig. 1b and Supplementary Fig. 2). We observed a strong reduction of Oct4-positive PGCs, specifically in female $Rnf2^{cko}$ embryos but not in male $Rnf2^{cko}$ or control embryos, starting around E12.5 of gestation (Fig. 1b, c). By contrast, double deficiency of *Ring1* and *Rnf2* caused a strong reduction of $Oct4^{+}$ PGCs by E11.5 in both sexes (Supplementary Fig. 6b, c), indicating an essential role for PRC1 in PGCs after their migration into the embryonic gonad (Supplementary Fig. 2)[11].

To further dissect the role of *Rnf2* in regulating Oct4 expression versus PGC development, we assessed co-expression of Oct4 and Rnf2 at E12.5 in $Cdh1^{+}$ PGCs in $Rnf2^{cko}$ embryos (Fig. 1d). The number of PGCs lacking detectable Rnf2 protein was strongly reduced in female but not male gonads, despite the fact that gonads of both sexes harboured comparable numbers of *Rnf2*-deficient PGCs at E12.0 (Supplementary Fig. 4a, and data not shown). We further observed a pronounced downregulation of Oct4 protein in female and some male *Rnf2*-deficient PGCs (Fig. 1d). These data indicate that *Rnf2* contributes to maintaining Oct4 expression, particularly in female PGCs, beginning between E12.0 and E12.5.

We subsequently investigated the mechanism underlying the reduction in female *Rnf2*-deficient PGCs. We failed to observe increased levels of apoptosis or major changes in cell cycle progression (data not shown). To study changes in gene expression, we introduced a green fluorescent protein (GFP) transgene driven by the promoter of *Oct4* lacking the proximal enhancer (*Oct4*(ΔPE)–GFP))[18] into the $Rnf2^{cko}$ strain and isolated pure populations of PGCs by fluorescence-activated cell sorting (FACS) (Fig. 1e). By quantitative reverse transcriptase PCR (qRT–PCR), we barely detected any *Rnf2* transcripts in isolated $Rnf2^{cko}$ PGCs, confirming efficient deletion of the $Rnf2^{fl}$ allele by *Tnap-cre* (Fig. 1f; see also Supplementary Fig. 4). Hence, we subsequently refer to PGCs from $Rnf2^{cko}$ embryos as $Rnf2^{\Delta}$. We also noticed significant reductions in GFP intensities in $Rnf2^{\Delta}$ PGCs isolated from male and female E11.5 and E12.5 embryos compared to controls (Fig. 1e), suggesting decreased *Oct4*(ΔPE) promoter activity in these cells. qRT–PCR analysis showed significantly reduced *Oct4* and *Nanog* expression in female GFP-positive $Rnf2^{\Delta}$ PGCs at E11.5 and E12.5 (Fig. 1f and data not shown). Thus, *Rnf2* is required for maintaining the expression of pluripotency factors in PGCs.

We next analysed genome-wide expression in purified PGCs. The number of misregulated genes in female compared to male $Rnf2^{\Delta}$ PGCs at E12.5 was 12-fold higher (Fig. 2a and Supplementary Table 1). Consistent with the fact that PRC1 is a transcriptional repressor, ~90% of misexpressed genes were upregulated. At E11.5, we only observed a few misregulated genes in $Rnf2^{\Delta}$ PGCs (Supplementary Table 1), consistent with the timing of the mutant phenotype. According to Gene Ontology analysis, gene functions related to meiosis (synapsis, sister chromatid cohesion) were highly over-represented among genes upregulated in female $Rnf2^{\Delta}$ PGCs. By contrast, nucleosome functions were enriched in downregulated genes, as reported previously for *Ring1* and *Rnf2* double-deficient germinal vesicle oocytes[10] (Supplementary Fig. 7 and Supplementary Table 2). Notably, we barely found any over-representation of developmental gene functions in $Rnf2^{\Delta}$ PGCs, a classical feature of PRC1 deficiency in ESCs[16,19] and germinal vesicle oocytes[10]. To test whether this is due to functional redundancy of the *Ring1* paralogue, we profiled expression in $Ring1^{\Delta}$ $Rnf2^{\Delta}$ PGCs purified from E11.5 embryos. Consistently, many developmental gene functions were over-represented among upregulated genes (Supplementary Fig. 8 and Supplementary Table 2). These data indicate that although *Ring1* expression is sufficient to safeguard global repression of canonical Polycomb target genes in PGCs, *Rnf2* is required in female E12.5 PGCs for repression of genes driving meiosis.

At E11.5, gonads still possess the potential to develop into either ovaries or testes, but are committed to a sex-specific differentiation process 2 days later. To relate aberrant expression in $Rnf2^{\Delta}$ PGCs to changes occurring during normal PGC differentiation, we profiled expression in $Rnf2^{+}$ PGCs isolated between E11.5 and E13.5. In this developmental period, sixfold more genes were upregulated in female (810) than in male (132) $Rnf2^{+}$ PGCs, with an additional 196 genes being upregulated in both sexes, suggesting the activation of female- and male-specific 'PGC-differentiation programs' (PDPs) (Fig. 2b). Among genes upregulated in E12.5 female $Rnf2^{\Delta}$ PGCs, 223 (43%) were found to be part of the female PDP (Fig. 2b and Supplementary Fig. 7b) (P value 2.67×10^{-159}; geometric test), being activated about 1 day ahead compared to those in $Rnf2^{+}$ PGCs (Fig. 2c).

©2013 Macmillan Publishers Limited. All rights reserved

We next identified that 24 out of 119 genes annotated with the Gene Ontology term 'meiosis' (GO:0007126) were part of the female PDP, probably reflecting activation of an 'early meiosis program' (Fig. 2d). Among these 24 genes, 10 were precociously upregulated in female $Rnf2^{\Delta}$ PGCs, including *Stra8*, required for meiotic initiation[5], and others such as *Rec8*, *Sycp3*, and *Hormad2*, which have key functions in cohesion, chromosome synapsis and recombination[20–22]. Using qRT–PCR, we measured high *Stra8*, *Rec8* and *Sycp3* messenger RNA levels at E13.5 in female $Rnf2^{+}$ PGCs (Fig. 3a). At E11.5 and E12.5, *Stra8* and *Rec8* were precociously activated, with up to 20-fold higher expression levels in female $Rnf2^{\Delta}$ versus $Rnf2^{+}$ PGCs. *Sycp3* was expressed in $Rnf2^{+}$ PGCs from both sexes with up to fourfold increased levels in female $Rnf2^{\Delta}$ PGCs (Fig. 3a). Immunofluorescence analyses revealed strong Stra8 nuclear localization by E12.5 in PGCs of $Rnf2^{cko}$ gonads, whereas the protein only started to accumulate in control PGCs at E13.5 (Fig. 3b and Supplementary Fig. 7c). For Sycp3, we observed focal nuclear staining at E13.5 and synaptonemal complex staining at E14.5 in $Rnf2^{\Delta}$ germ cells, whereas only axial elements of meiotic chromosomes were visible in $Rnf2^{+}$ germ cells at E14.5 (Fig. 3c, d, data not shown). These data indicate that precocious transcriptional activation of *Stra8* and other PGC differentiation and meiosis genes in female $Rnf2^{\Delta}$ PGCs induces these cells to prematurely stop proliferation and enter into meiotic prophase, hence accounting for the lower number of female $Rnf2^{\Delta}$ PGCs at E12.5. We also measured increased expression of genes functioning in retinoic acid metabolism, such as *Aldh1a2*, *Crabp1* and *Crabp2* in female $Rnf2^{\Delta}$ PGCs at E11.5 and 12.5 (Supplementary Fig. 7d and data not shown), probably enhancing retinoic acid signalling and meiotic entry in a feed-forward manner[23].

We next wanted to find out why *Stra8* transcription is abnormally activated only in female $Rnf2^{\Delta}$ PGCs. In male gonads, retinoic-acid-mediated induction of *Stra8* is counteracted by the somatically expressed retinoid-degrading enzyme Cyp26b1 and by the Fgf9 signalling pathway (Supplementary Fig. 2)[3,4,24]. PRC1 may therefore not be required to suppress *Stra8* expression in males. To study this possibility, we aimed to overcome the antagonizing activity of Cyp26b1 and cultured E11.5 genital ridges for 24 h in the presence of all-*trans* retinoic acid (ATRA)[3,4]. ATRA increased *Stra8* expression in isolated $Rnf2^{+}$ PGCs of both sexes to levels comparable with what we measured in female $Rnf2^{\Delta}$ PGCs treated with vehicle (Fig. 4a). Intriguingly, ATRA additionally enhanced *Stra8* expression in male as in female $Rnf2^{\Delta}$ PGCs to >fivefold higher levels compared to $Rnf2^{+}$ PGCs (Fig. 4a), indicating that PRC1 effectively suppresses retinoic-acid-induced *Stra8* activation in PGCs of both sexes. We observed comparable responses for *Rec8* and *Sycp3* expression (Fig. 4a). By contrast, *Stra8* activation was completely suppressed in all genotypes by treatment with Win-18446 (*N*,*N*′-octamethylenebis(dichloroacetamide)), an inhibitor of the retinoic acid biogenesis pathway[25] (Fig. 4a). Likewise, 2-day *in vivo* exposure of PGCs to Win-18446 in developing embryos suppressed premature Stra8 protein expression (Supplementary Fig. 9a, b). In addition, numbers of Oct4-expressing PGCs in $Rnf2^{cko}$ gonads were significantly increased upon treatment *in vitro* and *in vivo* with Win-18446 (Fig. 4b and Supplementary Fig. 9c, d). These results indicate that sensitization of PGCs to retinoic acid signalling by *Rnf2* deficiency contributes to precocious exit from the PGC state and activation of the meiotic program in female gonads.

Subsequently, most $Rnf2^{\Delta}$ PGCs were unable to complete meiosis and to develop into mature oocytes (Supplementary Fig. 10), possibly as a consequence of the precocious entry into meiosis that may eventually impair, perhaps in concert with other aberrant changes in gene expression (Fig. 2b), the execution of the natural female PDP and oogenesis. Alternatively, PRC1 may exert a separate essential function later during meiotic progression.

Finally, we performed a micro-chromatin immunoprecipitation (μChIP) assay on isolated $Rnf2^{+}$ PGCs. At E11.5, the *Stra8* promoter was strongly enriched with PRC2-mediated H3K27me3 and with

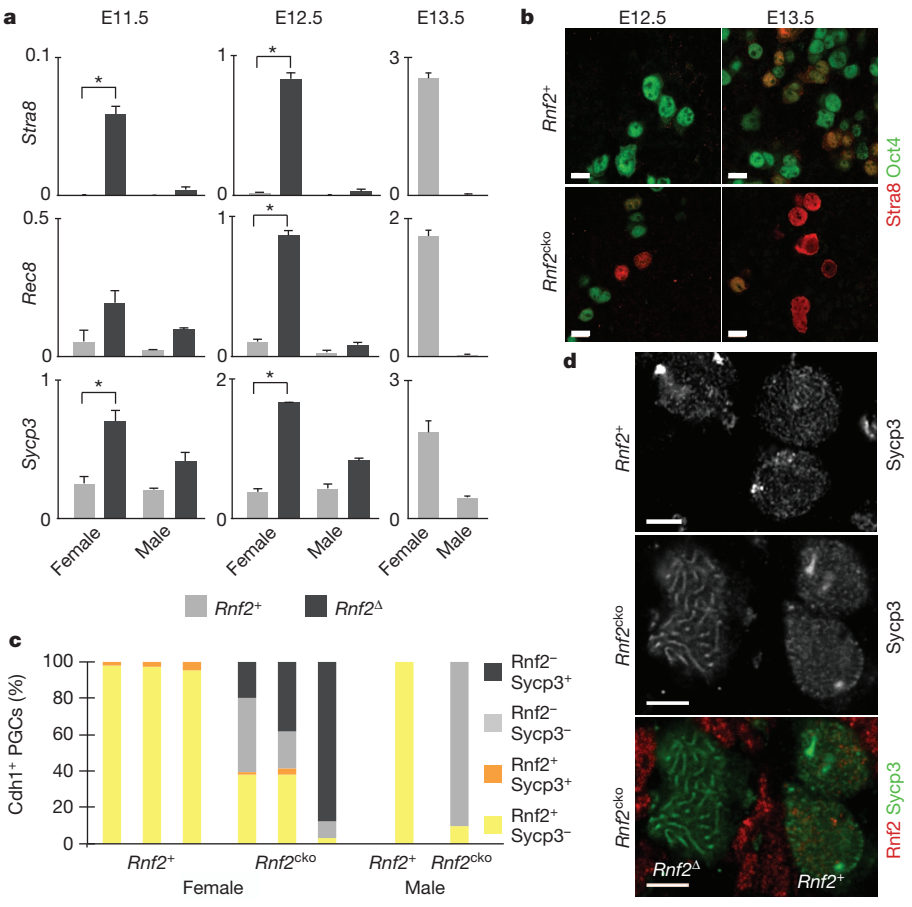


Figure 3 | Female $Rnf2^{\Delta}$ PGCs enter precociously into meiotic prophase.
a, Representative qRT–PCR data of *Stra8*, *Rec8 and Sycp3* transcripts (normalized to *Tbp*) in isolated PGCs. Error bars indicate +s.d. of two technical replicates. $*P < 0.05$ (Student's *t*-test).
b, Immunofluorescence staining of Oct4 and Stra8 in $Rnf2^{+}$ and $Rnf2^{cko}$ female gonads at E12.5 and E13.5. Scale bars, 10 μm. **c**, Classification of Cdh1-positive PGCs according to Rnf2 and Sycp3 protein levels in $Rnf2^{+}$ and $Rnf2^{cko}$ E13.5 gonads of individual embryos. Numbers in brackets indicate the number of PGCs scored per embryo.
d, Immunofluorescence staining of Sycp3 and Rnf2 in $Rnf2^{+}$ and $Rnf2^{cko}$ E14.5 ovaries. Scale bars, 10 μm.

©2013 Macmillan Publishers Limited. All rights reserved

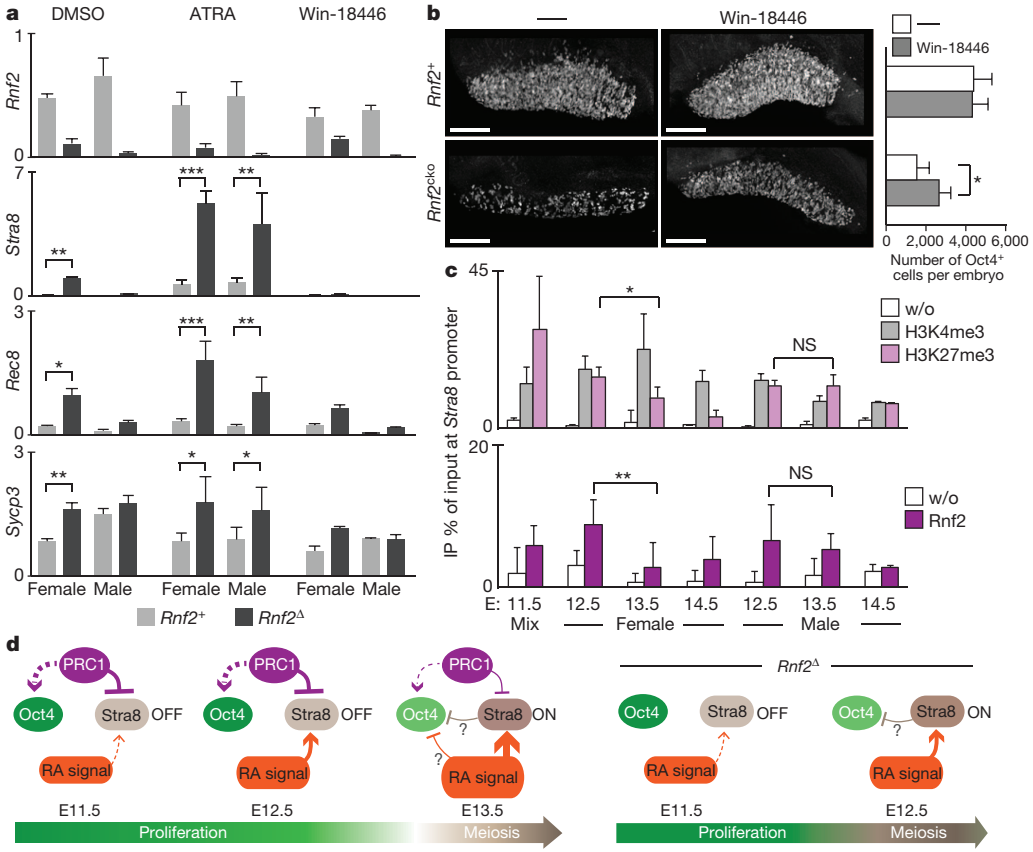


Figure 4 | PRC1 antagonizes retinoic acid signalling and maintains *Stra8* in a repressive chromatin state. **a**, Representative qRT–PCR data of *Rnf2*, *Stra8*, *Rec8* and *Sycp3* transcripts (normalized to *Tbp*) in isolated PGCs. Error bars indicate +s.d. of 2–4 technical replicates. $*P < 0.01$; $**P < 0.001$; $***P < 0.0001$ (Student's *t*-test). **b**, Immunofluorescence staining of Oct4 and number of Oct4-positive PGCs in $Rnf2^{+}$ and $Rnf2^{cko}$ whole gonads and mesonephroi from E12.5 female embryos, treated with Win-18446 or vehicle at E10.5. Graph shows means with +s.d. $n = 5$–6. Scale bars, 300 µm. **c**, µChIP analysis for H3K4me3, H2K27me3 and Rnf2 at the promoter region of *Stra8* in PGCs isolated from E11.5, 12.5, 13.5 and 14.5 $Rnf2^{+}$ gonads. Percentages of immunoprecipitated DNA (IP) compared to input are shown. w/o, without primary antibodies. Error bars indicate +s.d. of 1–3 biological and 2–4 technical replicates. $*P < 0.05$; $**P < 0.001$ (Student's *t*-test). **d**, Model for transcriptional regulation of meiotic genes (for example, *Stra8*) and pluripotency genes (for example, *Oct4*) during sex differentiation of female PGCs. RA, retinoic acid.

H3K4me3, indicative of a Polycomb repressed, yet potentially transcriptionally primed, chromatin state. During subsequent stages, we noticed a progressive decrease in H3K27me3 levels in female PGCs consistent with *Stra8* transcriptional activation, whereas the promoter remained bivalent in male PGCs (Fig. 4c). By contrast, the *Hoxa9* promoter was bivalent in all conditions (Supplementary Fig. 11). Likewise, we detected Rnf2 association with the *Stra8* promoter at E11.5 and E12.5 PGCs and a significant decrease in Rnf2 occupancy in female but not male PGCs at E13.5. Together, these experiments suggest that *Stra8* is directly regulated by PRC2 and PRC1 in PGCs of both sexes.

In summary, we identified an essential role for PRC1 in PGC development between E10.5 and E11.5. Later, between E11.5 and E13.5, *Rnf2* effectively modulates the sensitivity of PGCs to retinoic-acid-mediated induction of meiosis by directly controlling the competence of *Stra8* and probably of other meiotic genes for transcriptional activation. Like the proposed role of *Cbx2* in temporal co-linearity of *Hox* gene activation[26], we speculate that PRC1 maintains repression of *Stra8* and other genes of the PGC differentiation and early meiosis programs until retinoic acid signalling has reached a certain threshold (Fig. 4d). In addition, PRC1 regulates the expression of pluripotency genes in PGCs, possibly indirectly by suppressing transcription of negative regulators. Impairing PRC1 function in these parallel pathways probably leads to a synergistic effect, thereby promoting premature transition from proliferation into meiosis.

METHODS SUMMARY

Embryos were obtained by timed matings, by scoring noon of the day after mating as E0.5. Genital ridges were cultured in drops of Dulbecco's minimal eagle medium supplemented with 10% fetal calf serum at 37 °C with 5% CO_2 in air. For expression profiling, we collected in triplicate 500 PGCs per embryo by fluorescence-activated cell sorting and processed, hybridized to Affymetrix Mouse Gene 1.0 arrays and analysed as described previously[10]. Gene Ontology terms were obtained using GOstat (http://gostat.wehi.edu.au). µChIP was performed as described previously[27].

Full Methods and any associated references are available in the online version of the paper.

Received 27 April 2012; accepted 16 January 2013.

1. Brennan, J. & Capel, B. One tissue, two fates: molecular genetic events that underlie testis versus ovary development. *Nature Rev. Genet.* **5,** 509–521 (2004).
2. Menke, D. B., Koubova, J. & Page, D. C. Sexual differentiation of germ cells in XX mouse gonads occurs in an anterior-to-posterior wave. *Dev. Biol.* **262,** 303–312 (2003).
3. Koubova, J. *et al.* Retinoic acid regulates sex-specific timing of meiotic initiation in mice. *Proc. Natl Acad. Sci. USA* **103,** 2474–2479 (2006).
4. Bowles, J. *et al.* Retinoid signaling determines germ cell fate in mice. *Science* **312,** 596–600 (2006).
5. Baltus, A. E. *et al.* In germ cells of mouse embryonic ovaries, the decision to enter meiosis precedes premeiotic DNA replication. *Nature Genet.* **38,** 1430–1434 (2006).
6. Deschamps, J. Ancestral and recently recruited global control of the *Hox* genes in development. *Curr. Opin. Genet. Dev.* **17,** 422–427 (2007).
7. Duester, G. Retinoic acid synthesis and signaling during early organogenesis. *Cell* **134,** 921–931 (2008).
8. Oulad-Abdelghani, M. *et al.* Characterization of a premeiotic germ cell-specific cytoplasmic protein encoded by Stra8, a novel retinoic acid-responsive gene. *J. Cell Biol.* **135,** 469–477 (1996).
9. Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nature Rev. Cancer* **6,** 846–856 (2006).
10. Posfai, E. *et al.* Polycomb function during oogenesis is required for mouse embryonic development. *Genes Dev.* **26,** 920–932 (2012).
11. Saitou, M., Kagiwada, S. & Kurimoto, K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* **139,** 15–31 (2012).
12. Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature Rev. Mol. Cell Biol.* **10,** 697–708 (2009).
13. Di Carlo, A. & De Felici, M. A role for E-cadherin in mouse primordial germ cell development. *Dev. Biol.* **226,** 209–219 (2000).
14. Puschendorf, M. *et al.* PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nature Genet.* **40,** 411–420 (2008).
15. Lomelí, H., Ramos-Mejia, V., Gertsenstein, M., Lobe, C. G. & Nagy, A. Targeted insertion of Cre recombinase into the TNAP gene: excision in primordial germ cells. *Genesis* **26,** 116–117 (2000).
16. Endoh, M. *et al.* Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135,** 1513–1524 (2008).

©2013 Macmillan Publishers Limited. All rights reserved

17. del Mar Lorente, M. *et al.* Loss- and gain-of-function mutations show a polycomb group function for Ring1A in mice. *Development* **127,** 5093–5100 (2000).
18. Yoshimizu, T. *et al.* Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev. Growth Differ.* **41,** 675–684 (1999).
19. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* **24,** 265–276 (2010).
20. Xu, H., Beasley, M. D., Warren, W. D., van der Horst, G. T. & McKay, M. J. Absence of mouse REC8 cohesin promotes synapsis of sister chromatids in meiosis. *Dev. Cell* **8,** 949–961 (2005).
21. Yuan, L. *et al.* Female germ cell aneuploidy and embryo death in mice lacking the meiosis-specific protein SCP3. *Science* **296,** 1115–1118 (2002).
22. Wojtasz, L. *et al.* Meiotic DNA double-strand breaks and chromosome asynapsis in mice are monitored by distinct HORMAD2-independent and -dependent mechanisms. *Genes Dev.* **26,** 958–973 (2012).
23. Maden, M. Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nature Rev. Neurosci.* **8,** 755–765 (2007).
24. Bowles, J. *et al.* FGF9 suppresses meiosis and promotes male germ cell fate in mice. *Dev. Cell* **19,** 440–449 (2010).
25. Hogarth, C. A. *et al.* Suppression of *Stra8* expression in the mouse gonad by WIN 18,446. *Biol. Reprod.* **84,** 957–965 (2011).
26. Bel-Vialar, S. *et al.* Altered retinoic acid sensitivity and temporal expression of *Hox* genes in *polycomb-M33*-deficient mice. *Dev. Biol.* **224,** 238–249 (2000).
27. Dahl, J. A. & Collas, P. A rapid micro chromatin immunoprecipitation assay (ChIP). *Nature Protocols* **3,** 1032–1045 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Nagy and H. Schöler for the *Tnap-cre* and *Oct4*(*ΔPE*)–GFP mice, respectively, and M. Griswold for the Stra8 antibody. We are grateful to L. Gelman (microscopy and imaging), L. Burger, M. Stadler (bioinformatics), E. Cabuy, S. Thiry, K. Jacobeit (functional genomics) and the FMI animal facility for assistance. We thank members of the Peters laboratory, particularly M. Tardat, S. Erkek and M. Gill, for experimental support and discussions. S.Y. is a recipient of a Human Frontier Science Program long-term fellowship and a Japan Society for the Promotion of Science postdoctoral fellowship. Research in the Peters laboratory is supported by the Novartis Research Foundation, the Swiss National Science Foundation (31003A_125386 and NRP 63 - Stem Cells and Regenerative Medicine), SystemsX.ch (Cell plasticity), the Japanese Swiss Science and Technology Cooperation Program, the European Network of Excellence 'The Epigenome' and the EMBO Young Investigator program.

Author Contributions S.Y. and A.H.F.M.P. conceived and designed the experiments. S.Y. performed almost all experiments. C.-Y.L. performed μChIP experiments. H.K. performed FACS isolations. Z.L. isolated germinal vesicle oocytes. P.N. isolated RNA for RNA sequencing and provided advice on specific target amplification qRT-PCR. M.V. provided *Ring1*-deficient mice. M.v.L. provided *Rnf2* conditionally deficient mice. T.C.R. assisted in microarray and RNA-sequencing analysis. S.Y., C.-Y.L. and A.H.F.M.P. analysed the data. S.Y. and A.H.F.M.P. wrote the manuscript.

Author Information Microarray and RNA-sequencing data have been deposited in the Gene Expression Omnibus under accession numbers GSE42782 and GSE42852, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.H.F.M.P. (antoine.peters@fmi.ch).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Mice and embryo collection. *Rnf2*^{cko} mice with *Rnf2*-deficient PGCs were generated by combining a floxed *Rnf2* (*Rnf2*^{fl}) allele[14] with the *Tnap-cre* transgene allele as illustrated in Supplementary Fig. 3. Introduction of the *Ring1* mutation[17] is illustrated in Supplementary Fig. 6a. For PGC isolation, embryos were sired by fathers that were homozygous for the *Oct4*(*ΔPE*)–GFP transgene. Mice were maintained on a mixed background of 129/Sv and C57BL/6J. Embryos were obtained by timed matings, by scoring noon of the day following mating as 0.5 embryonic day of development (E0.5). The genotype of embryos was determined by PCR as described previously[14]. Sex of gonads was determined by PCR for *Ubex1* using the following primers: forward, 5′-TGGTCTGGACCCAAACGCTG TCCACA-3′; reverse, 5′- GGCAGCAGCCATCACATAATCCAGATG-3′. All experiments were performed in accordance with the Swiss Animal Protection laws and institutional guidelines.

Antibodies. For immunofluorescence analyses, the following primary and secondary antibodies were used: polyclonal anti-Oct4 (sc-8628, 1:150), monoclonal anti-Rnf2 (gift from H. Koseki, 1:400)[28], monoclonal anti-E-cadherin (Invitrogen, 1:250), polyclonal anti-Stra8 (rabbit, gift from M. Griswold, 1:1,000)[29], polyclonal anti-Sycp3 (rabbit, gift from C. Heyting, 1:500)[30], polyclonal anti-Mel18 (sc-10774, 1:50), polyclonal anti-Rybp (1:400)[31], monoclonal anti-H2AK119u1 (Cell signaling, 1:500), polyclonal anti-H3K27me3 (gift from T. Jenuwein, 1:500)[32], monoclonal anti-Ezh2 (Novocastra, 1:200), anti-goat IgG-Alexa 488, anti-rabbit IgG-Alexa 488, anti-mouse IgG-Alexa 555 and anti-rat Cy5. For ChIP analysis, anti-H3K4me3 (Millipore, 17-614), anti-H3K27me3 (Millipore, 07-449) and anti-Rnf2 (Active motif, 39663) were used.

Immunofluorescence. For whole-mount stainings, dissected gonads with mesonephroi were fixed for 15 min in 3% paraformaldehyde in PBS (pH 7.4) and permeabilized with 0.5% Triton X-100 in PBS for 20 min on ice. Fixed embryos were blocked overnight at 4 °C in PBS containing 0.1% Triton X-100, 10% BSA and 5% normal donkey serum, and were then incubated with primary antibodies in blocking solution overnight at 4 °C. Gonads were washed three times for 1 h in PBS containing 0.1% Triton X-100 and 2% BSA before application of secondary antibodies. For detection, secondary antibodies were diluted 1:500 in blocking solution and gonads were incubated overnight at 4 °C followed by three washing steps for 1 h in PBS with 0.1% Triton X-100. Gonads were stained briefly with DAPI and mounted in Vectashield (Vector). For gonadal section stainings, the posterior part of embryos or gonads with mesonephroi were frozen in Tissue-Tek Optimal Cutting Temperature (OCT) compound (Sakura Finetek) on dry ice. Alternatively, the materials were fixed with 3% paraformaldehyde in PBS (pH 7.4) for 10 min, soaked in 30% sucrose solution overnight and embedded in OCT compound. Twelve-micron-thick cryo-sections were cut from frozen blocks with Microm HM355S. Cryo-sections were fixed with 3% paraformaldehyde for 10 min at room temperature (about 20 °C), permeabilized in 0.5% Triton X-100 in PBS for 4 min at 4 °C and blocked for 30 min in PBS containing 1% BSA at room temperature. Sections were incubated with primary antibodies in blocking solution overnight at 4 °C and subsequently washed three times for 10 min in PBS with 0.05% Tween-20. Incubation of secondary antibodies was done in the blocking solution for 1 h at room temperature.

Microscopy and image analysis. Immunofluorescence stainings of gonads were analysed using the Zeiss LSM 700 confocal microscope. For whole-mount gonads, images were acquired by using a tile function with a *z*-series of 1-µm slices in ZEN software and whole image was reconstructed using the XUV-tools software. We counted the number of Oct4-positive cells using a spot function in Imaris (Bitplane) software.

Isolation of PGCs expressing the *Oct4*(*ΔPE*)–GFP transgene by FACS. Dissected gonads were enzymatically disrupted using 0.025% trypsin at 37 °C for 8 min. Trypsin activity was inhibited by adding fetal calf serum in Hank's buffer salt solution without phenol red. Gonads were dispersed by pipetting and subjected to FACS. Embryos were processed individually for expression analysis. PGCs isolated from several embryos were pooled for ChIP analysis.

qRT–PCR. Total RNA was extracted from isolated PGCs or surrounding somatic cells from individual embryos using the PicoPure RNA Isolation Kit (KIT0202) according to the manufacturer's instructions (Stratagene) with the addition of 100 ng *Escherichia coli* ribosomal RNA as carrier. RT–PCR was performed using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol. qPCR reactions were performed with complementary DNA corresponding to 20 cells using the SYBR Green PCR Master Mix (Applied Biosystem) in an ABI Prism 7000 Real time PCR machine. All qPCR measurements were normalized to the endogenous expression level of *Tbp*. We performed qRT–PCR analyses on multiple pairs of *Rnf2*^{cko} and control littermates. Data in figures present technical replicates for pairs of genotypes. Primers used were as follows: *Rnf2* (forward, 5′-TTAGAAGTGGCAACAAAGAGTG-3′; reverse, 5′-CGCTTCATACTCATCACGAC-3′), *Oct4* (forward, 5′-GATGCTGTGAGCCA AGGCAAG-3′; reverse, 5′-GGCTCCTGATCAACAGCATCAC-3′), *Nanog* (forward, 5′-CTTTCACCTATTAAGGTGCTTGC-3′; reverse, 5′-TGGCATCGG TTCATCATGGTAC-3′), *Stra8* (forward, 5′-CAAAAGCCTTGGCTGTGTTA-3′; reverse, 5′-AAAGGTCTCCAGGCACTTCA-3′), *Rec8* (forward, 5′-CCAA CAAGGAGCTGGACTTC-3′; reverse, 5′-GGACAGCACCAAGAGCAGAT-3′), *Sycp3* (forward, 5′-GTGTTGCAGCAGTGGGAAC-3′; reverse, 5′-GCTTT CATTCTCTGGCTCTGA-3′), *Aldh1a2* (forward, 5′-CCCTGACAGTGGCT TTGAGT-3′; reverse, 5′-CTGTGGGTTGAAGGGAGCTA-3′), *Crabp1* (forward, 5′-GCTTCGAGGAGGAGACAGTG-3′; reverse, 5′- CAGCTCTCGGGTCCAG TAAG-3′), *Crabp2* (forward, 5′-GCCGAGAACTGACCAATGAT-3′; reverse, 5′-GGAAGTCGTCTCAGGCAGTT-3′), *Tbp* (forward, 5′-TGCTGTTGGTGAT TGTTGGT-3′; reverse, 5′-AACTGGCTTGTGTGGGAAAG-3′).

Expression profiling of PGCs and data analysis. We performed expression profiling on PGCs isolated from three pairs of *Rnf2*^{cko} and control littermates for each developmental time point. RNA was extracted from 500 PGCs isolated per embryo using the PicoPure RNA Isolation Kit (KIT0202) according to the manufacturer's instructions (Stratagene). The quality of the RNA was assessed using the Agilent 2100 Bioanalyzer and RNA 6000 Pico Chip. The extracted RNA was converted into OmniPlex Whole transcriptome amplification (WTA) cDNA libraries and amplified by WTA PCR using reagents supplied with the TransPlex WTA1 kit (Sigma) following the manufacturer's instructions with minor modifications. The obtained cDNA was purified using the GeneChip cDNA Sample Cleanup Module (Affymetrix). The labelling, fragmentation and hybridization of cDNA was performed according to Affymetrix instructions (GeneChip Whole Transcription Sense Target Labelling technical manual, Rev. 2) with minor modifications. Samples were hybridized to Affymetrx Mouse Gene 1.0 arrays. Microarray quality control and analysis was carried out in R 2.10.0 and Bioconductor 2.5. In brief, array quality was assessed using the 'arrayQualityMetrics' package. Raw data was read into R and normalized with RMA using the 'affy' package and differentially expressed genes were identified using the empirical Bayes method (*F* test) implemented in the LIMMA package. *P* values were adjusted for false discovery rate (FDR) using the Benjamini and Hochberg correction. Probe sets with a log2 average contrast signal of at least 3, an adjusted *P* value of <0.05, and an absolute linear fold-change of at least 1.5-fold were selected. The *P* values reported for enriched Gene Ontology terms were obtained using GOstat (http://gostat.wehi.edu.au) (Supplementary Figs 7 and 8 and Supplementary Table 2). A list of genes belonging to the GO term 'meiosis' (GO:0007126) were obtained using R annotation packages (library(org.Mm.eg.db) and library(GO.db)) (Fig. 2d and Supplementary Table 1). For RNA-sequencing experiments, RNA was extracted from 500 PGCs per embryo using RNeasy Micro Kit (Qiagen). The RNA amplification and cDNA generation were performed using NuGEN Ovation RNA-seq System V2 (Part no. 7102) and sequencing libraries were prepared using TruSeq DNA Sample Preparation Kit (low-throughput protocol) (Part no. 15005180 Rev. C). Barcoded libraries were sequenced in one lane of Illumina HiSeq instrument. The resulting sequencing reads were filtered, aligned to Refseq gene models and weighted as described previously[33].

Treatment of gonads with agonist and antagonist of retinoic acid signalling. Both gonads per embryo were dissected and were cultured together or separately in a drop of Dulbecco's minimal eagle medium supplemented with 10% fetal calf serum at 37 °C in a humidified atmosphere of 5% CO_2 in air. ATRA (Sigma) and Win-18446 (ABCR) were dissolved in dimethylsulphoxide (DMSO). These compounds were added to culture media with concentration of 0.5 µM for ATRA or 2 µM for Win-18446. Control cultures were treated with DMSO vehicle as appropriate. For *in vivo* administration of Win-18446, 100 mg per ml stock solution (in DMSO) was diluted with oil and 2.5 mg was injected intraperitoneally into pregnant female mice at E10.5. Embryos were collected at E12.5. DMSO mixed with oil was injected into pregnant female mice for control experiments.

Specific target amplification (STA) qRT–PCR on single germinal vesicle oocytes. Germinal vesicle oocytes were isolated from ovaries and carefully washed with removing cumulus cells. STA was performed using the CellsDirect One-Step qRT–PCR Kit (11753-100) according to the manufacturer's instructions (Invitrogen). In brief, the oocytes were individually added to the reaction mix containing RNase Inhibitor (Ambion) and primers of target genes of interest. After amplification, the samples were treated with Exonuclease I (New England Biolabs) and used as template for qPCR. Primers used were as follows: *Rnf2* (forward, 5′-CAGGCCCCATCCAACTCTTA-3′; reverse, 5′-CAACAGTGGCA TTGCCTGAA-3′), *Ssu72* (forward, 5′-GGTGTGCTCGAGTAACCAGAA-3′; reverse, 5′- CAAAGGAGCGGACACTGAAAC-3′).

µChIP analysis. Small-scale ChIP experiment was performed as previously described[27] with some modifications. In brief, 15,000 FACS-sorted PGCs were cross-linked with 0.5% paraformaldehyde in PBS for 10 min at room temperature and quenched with 125 mM glycine. PGCs were then lysed in 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS, and protease inhibitors. The cell lysate was sonicated in a

©2013 Macmillan Publishers Limited. All rights reserved

Diagenode Bioruptor to achieve a mean DNA fragment size of around 200–400 base pairs. After centrifugation, the supernatants were diluted with radio immunoprecipitation assay (RIPA) buffer to an equivalent of 2,500 PGCs and incubated with antibody-bound protein G-magnetic beads overnight at 4 °C. The beads were washed four times with the RIPA buffer and one time with 10 mM Tris HCl, pH 8.0, 10 mM EDTA buffer, and bead-bound complexes were incubated with complete elution buffer (20mM Tris-HCl, pH 7.5, 5 mM EDTA, 50 mM NaCl, 1% SDS, 50 mg per ml proteinase K) at 68 °C for DNA elution, crosslink reversal and protein digestion. Finally, immunoprecipitated DNA was purified by phenol-chloroform extraction and ethanol precipitation and dissolved in MilliQ water for qRT–PCR analysis. Primers used were as follows: *Stra8* (forward, 5′-GTATCGCC GTAACTCCCAGA-3′; reverse, 5′-GCAGATGACCCTCACACAAG-3′), *HoxA9* (forward, 5′- GGAGGGAGGGGAGTAACAAA-3′; reverse, 5′-TCACCTCGCCT AGTTTCTGG-3′).

28. Atsuta, T. *et al.* Production of monoclonal antibodies against mammalian Ring1B proteins. *Hybridoma* **20,** 43–46 (2001).
29. Zhou, Q. *et al.* Expression of stimulated by retinoic acid gene 8 (Stra8) and maturation of murine gonocytes and spermatogonia induced by retinoic acid in vitro. *Biol. Reprod.* http://dx.doi.org10.1095/biolreprod.107.064337 (2008).
30. Lammers, J. H. *et al.* The gene encoding a major component of the lateral elements of synaptonemal complexes of the rat is related to X-linked lymphocyte-regulated genes. *Mol. Cell. Biol.* **14,** 1137–1146 (1994).
31. García, E., Marcos-Gutierrez, C., del Mar Lorente, M., Moreno, J. C. & Vidal, M. RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. *EMBO J.* **18,** 3404–3418 (1999).
32. Peters, A. H. *et al.* Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12,** 1577–1589 (2003).
33. Tippmann, S. C. *et al.* Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels. *Mol. Syst. Biol.* **8,** 593 (2012).

©2013 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature11979

Ovarian surface epithelium at the junction area contains a cancer-prone stem cell niche

Andrea Flesken-Nikitin[1], Chang-Il Hwang[1], Chieh-Yang Cheng[1], Tatyana V. Michurina[2,3], Grigori Enikolopov[2,3] & Alexander Yu. Nikitin[1]

Epithelial ovarian cancer (EOC) is the fifth leading cause of cancer deaths among women in the United States, but its pathogenesis is poorly understood[1–3]. Some epithelial cancers are known to occur in transitional zones between two types of epithelium, whereas others have been shown to originate in epithelial tissue stem cells[4–6]. The stem cell niche of the ovarian surface epithelium (OSE), which is ruptured and regenerates during ovulation, has not yet been defined unequivocally. Here we identify the hilum region of the mouse ovary, the transitional (or junction) area between the OSE, mesothelium and tubal (oviductal) epithelium, as a previously unrecognized stem cell niche of the OSE. We find that cells of the hilum OSE are cycling slowly and express stem and/or progenitor cell markers ALDH1, LGR5, LEF1, CD133 and CK6B. These cells display long-term stem cell properties *ex vivo* and *in vivo*, as shown by our serial sphere generation and long-term lineage-tracing assays. Importantly, the hilum cells show increased transformation potential after inactivation of tumour suppressor genes *Trp53* and *Rb1*, whose pathways are altered frequently in the most aggressive and common type of human EOC, high-grade serous adenocarcinoma[7,8]. Our study supports experimentally the idea that susceptibility of transitional zones to malignant transformation may be explained by the presence of stem cell niches in those areas. Identification of a stem cell niche for the OSE may have important implications for understanding EOC pathogenesis.

An extensive integrated genomic analysis of 489 high-grade serous ovarian adenocarcinomas has provided important insights into the repertoire of molecular aberrations that are characteristic for this most common and deadly form of EOC[7]. However, interpretation of such results is complicated by continuous controversy with regard to the cell of origin of this disease. The OSE, the fimbrial epithelium of the uterine (Fallopian) tubes, and other derivatives of the secondary Mullerian system have been proposed as likely sources of EOC based on pathological observations, genetic and immunohistochemical studies of precursor lesions in human cancer and experimental approaches in several species[2,3,9,10]. Recently, it has been proposed that transitional (or junction) regions between the OSE, mesothelium and tubal epithelium may have more plastic and, presumably, less differentiated states. If this is the case, such regions would be candidate locations of origin of EOC[5,11,12]. However, so far there has been no direct experimental data to support this idea.

Stem cells, such as those of blood and hair, are slow-cycling and retain DNA labels in pulse-chase experiments. OSE cells with some stem and/or progenitor (stem/progenitor) cell properties have been identified previously based on their slow proliferation in label-retention assays[13]. Unfortunately, it remains uncertain whether these cells have potential for long-term self-renewal, a key feature of stem cells. It is also unknown whether these cells occupy anatomically defined areas, similar to those in other organs, such as the intestine, hair follicle, cornea and prostate[14,15]. Activity of the members of the detoxifying enzyme aldehyde dehydrogenase (ALDH) family and the corresponding expression of ALDH1 have been identified as useful markers of stem/progenitor cells in a number of cell lineages, such as mammary, prostate, colon, haematopoietic, neural and mesenchymal cell lineages[16–19]. It has been also reported that approximately 7.6% of OSE cells have high ALDH activity[20]. We therefore tested whether detection of this activity can be used to enrich normal OSE stem/progenitor cells. We separated primary OSE cells into populations with high ALDH expression ($ALDH^+$) and those with low ALDH expression ($ALDH^-$), according to their level of ALDEFLUOR fluorescence by fluorescence-activated cell sorting (FACS, Supplementary Fig. 1), and subjected them to newly established monoclonal OSE-sphere formation assays (Fig. 1a, b and Supplementary Figs 2 and 3). $ALDH^+$ cells represented 5.3% of the OSE population, and they formed larger OSE spheres and at significantly higher frequency compared to $ALDH^-$ cells (Supplementary Fig. 4 and Supplementary Table 1). Consistent with expected self-renewal properties of stem/progenitor cells, spheres were formed from a single-cell suspension in at least five consecutive rounds of sphere dissociation and regeneration (Fig. 1c).

To identify $ALDH1^+$ OSE cells in the mouse ovary we carried out immunohistochemical staining for ALDH1 expression in ovarian tissues from 3-, 6- and 8-week-old virgin FVB/N, C57BL/6J, B6;129 and Swiss Webster mice. As the ovary has well-defined anatomical regions, we evaluated the expression of ALDH1 in the OSE covering the corpus luteum, the distal, antral and the hilum regions, as well as in the tubal epithelium. Notably, in mice of all age and strain groups, cells with high ALDH1 activity were detected predominantly in the hilum region (Fig. 1d and Supplementary Fig. 5). This region represents the point at which nerves and vessels enter the ovary and is covered by the epithelium representing the transition between OSE, mesothelium and tubal epithelium.

To determine the location of slow proliferating cells we carried out pulse-chase experiments by injecting 6- to 7-week-old virgin FVB/N mice for 10 days with a BrdU (5-bromodeoxyuridine) solution, after which we carried out detection of label-retaining cells (LRCs) immediately and monthly for up to 3 months after the pulse (Supplementary Fig. 6a). Ovarian regional analysis has revealed that the hilum contains the highest percentage of slow-cycling OSE cells by 3 months after the BrdU pulse (Fig. 1e and Supplementary Fig. 6b). These LRCs expressed ALDH1, as shown by double immunofluorescence (Fig. 1f). In agreement with earlier reports[13,21], proliferation of OSE was significantly greater after ovulation. To test whether LRCs preferentially divide after ovulation, cells that were double-positive for BrdU and Ki67 were scored in the hilum areas of mice that had been subjected to 2 and 3 months of BrdU pulse-chase experiments and that were euthanized in pro-oestrus (before ovulation) or in oestrus (after ovulation). Double-positive cells were present in the hilum OSE of mice in oestrus only, no such cells were found in pro-oestrus mice (Supplementary Fig. 7), suggesting that these LRCs were activated to repair the ruptured OSE.

We assessed the growth potential of OSE cells isolated by microdissection from the hilum (posterior) and opposite (anterior) regions

[1]Department of Biomedical Sciences and Cornell Stem Cell Program, Cornell University, Ithaca, New York 14853, USA. [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. [3]NBIC, Moscow Institute of Physics and Technology, 123182 Moscow, Russia.

©2013 Macmillan Publishers Limited. All rights reserved

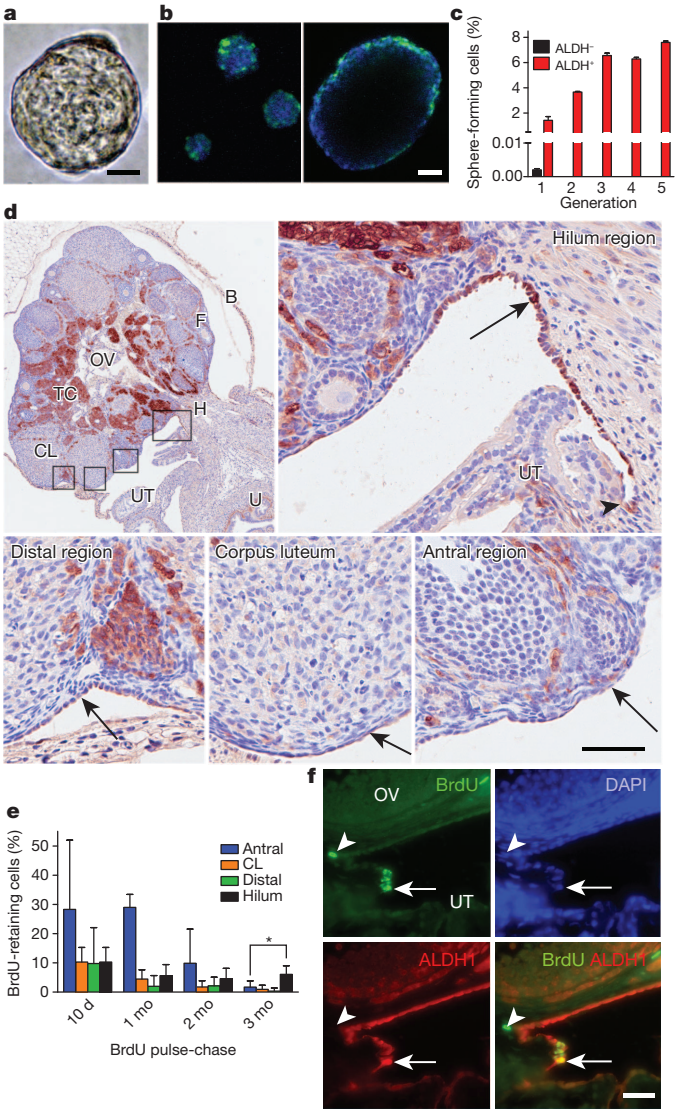


Figure 1 | Identification and location of putative OSE stem cells. **a**, OSE sphere, phase contrast. Scale bar, 20 µm. **b**, Confocal imaging of compact (left) and expanded (right) OSE spheres from *ß-actin-eGFP* mice at 4 and 10 days after placing in culture, respectively. Green fluorescence counterstaining with DAPI (4′,6-diamidino-2-phenylindole), blue. Scale bar, 50 µm. **c**, Frequency of sphere formation by ALDH$^-$ and ALDH$^+$ OSE cells for five consecutive generations (G, dissociation and clonal formation; $n = 6$). ALDH$^-$-derived cells very rarely formed spheres in G2 and did not yield any spheres in G3. **d**, Tissue sections from an ovary of a 6-week-old mouse, stained for ALDH1 expression. ALDH1 (brown) is expressed preferentially in the OSE (arrows) of the hilum region (H) as compared to that of the antral region, corpus luteum (CL) or distal region. ALDH1 staining is also present in the theca cells (TC) of the ovary (OV). Rectangles in the top-left image indicate respective locations (clockwise) of the regions in the mouse ovary. The junction between OSE and tubal epithelium is indicated by an arrowhead. B, bursa; F, follicle; U, uterus; UT, uterine tube. Elite ABC method, haematoxylin counterstaining. Scale bar represents 500 µm (top-left image) or 50 µm (all other images). **e**, Quantification of BrdU label-retaining cells (LRCs) in the antral, corpus luteum (CL), distal and hilum regions ($n = 4$). At 3 months (mo) after BrdU pulse, the percentage of BrdU-retaining cells was significantly higher in the hilum versus antral region (*$P = 0.0005$, two-tailed unpaired *t*-test), versus corpus luteum or distal region ($P < 0.0001$, two-tailed unpaired *t*-test). **f**, Detection of BrdU LRCs (green), ALDH1 (red) and overlay (orange) in the hilum OSE (arrows) after 3 months of chase. Arrowhead, BrdU LRC in the neighbouring stroma. Counterstaining with DAPI, blue. Scale bar, 50 µm. All error bars denote s.d.

of the ovary (Supplementary Fig. 8). After 3 days of primary culture, hilum OSE cells formed significantly more large colonies (≥20 cells) than OSE cells from the anterior region (Fig. 2a, b). Notably, compared to other OSE cells, hilum cells developed spheres at significantly higher frequency and could be propagated for at least 7 generations (Fig. 2c, Supplementary Fig. 9 and Supplementary Table 2). These findings are consistent with results of our previous experiments with ALDH$^+$ cells.

For additional phenotypical characterization, ALDH$^+$ and ALDH$^-$ OSE cells were isolated by FACS and their RNA used for gene-expression profiling (Fig. 2d, Supplementary Fig. 10 and Supplementary Table 3). *Aldh1* gene was among the most highly expressed genes in ALDH$^+$ cells (Supplementary Fig. 10). LGR5, CD133 (also known as prominin 1), CK6B and LEF1 were consistently more highly expressed than other known stem cell markers in ALDH$^+$ cells (Fig. 2e). Expression of these markers in the hilum cells was also confirmed by immunostaining (Fig. 2f). Consistent with this, we have found that some of the

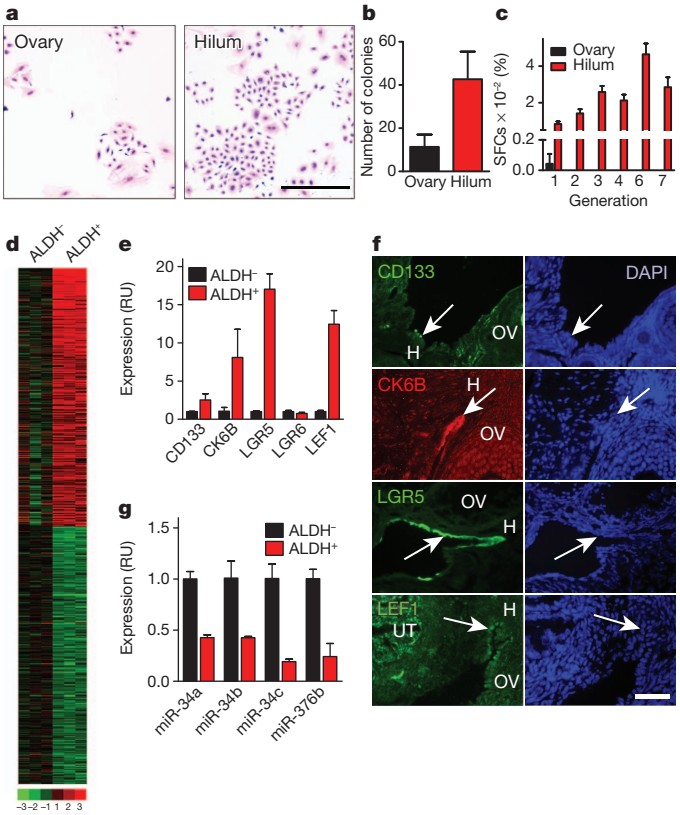


Figure 2 | Functional characterization of the hilum OSE cells. **a**, Colony formation by OSE cells isolated from the anterior region of the ovary and the hilum, shown by Giemsa staining. 1×10^4 cells were cultured per plate. Scale bar, 500 µm. **b**, Quantitative analysis of the frequency of large colonies (≥20 cells) formed by OSE cells from the anterior part ($n = 8$) and the hilum ($n = 6$). $P < 0.0001$, two-tailed unpaired *t*-test. **c**, Frequency of OSE sphere-forming cells (SFCs) in the anterior region of the ovary and hilum for seven consecutive generations (G, $n = 3$). Cells derived from the anterior region of the ovary very rarely formed spheres in G2 and did not produce any spheres in G3. **d**, Gene-expression profiles of 3 independent pools ($n = 10$ mice each) of ALDH$^-$ and ALDH$^+$ cells. The keys shows fluorescence expression ratios ($\log_2$ transformed). **e**, Expression of stem cell markers in ALDH$^-$ and ALDH$^+$ cells. Quantitative real-time polymerase chain reaction (rtPCR) ($n = 3$, $P < 0.01$ in all cases, except for LGR6). RU, relative units. **f**, Detection of hilum cells (arrows) expressing CD133, CK6B, LGR5 and LEF1. Immunofluorescence (CD133, CK6B, and LEF1) or eGFP expression under the control of the *Lgr5* promoter in *Lgr5-eGFP-IRES-CreERT2* mice. Counterstaining with DAPI, blue. Scale bar, 50 µm. **g**, Expression of microRNAs in ALDH$^-$ and ALDH$^+$ cells. Quantitative rtPCR ($n = 3$, $P < 0.01$ in all cases). All error bars denote s.d.

©2013 Macmillan Publishers Limited. All rights reserved

microRNAs that counteract stem cell properties, such as microRNAs of the miR-34 family[22–24], and miR-376b (Palmaccio, S. J., Cheng, L., A.F.-N. and A.Yu.N., unpublished observations), are preferentially downregulated in ALDH$^+$ OSE cells relative to the ALDH$^-$ OSE population (Fig. 2g).

As hilum cells express LGR5, to trace the fate of these cells we used *Lgr5*$^{tm1(cre/ERT2)Cle}$*/J* mice, which harbour an *Lgr5-eGFP-IRES-CreERT2* knock-in allele[25]. In these mice, enhanced green fluorescent protein (eGFP) and Cre–ERT2 are expressed in LGR5-positive cells. However, activity of Cre–ERT2 fusion protein occurs only after induction by tamoxifen. We crossed *Lgr5*$^{tm1(cre/ERT2)Cle}$*/J* mice with *Gt(ROSA)26Sor*$^{tm9(CAG\text{-}tdTomato)Hze}$*/J* (Ai9) mice (Supplementary Fig. 11). In Ai9 mice, expression of red fluorescent protein variant (tdTomato) under the control of *CAG* promoter at the *Rosa26* locus is possible only after Cre-mediated deletion of the stop codon flanked by *loxP* sites[26]. To test whether *Lgr5* promoter directs Cre expression to the hilum cells, mice were exposed to a single dose of tamoxifen and their ovaries were collected 1 and 3 days later. Microscopy analysis showed that cells of the hilum were exclusively labelled in the OSE at these early time points (Fig. 3a–c and Supplementary Fig. 12a–g). Control experiments included administration of oil to double-knock-in littermates (Fig. 3d, e and Supplementary Fig. 12h–j). Furthermore, wild-type mice and mice carrying only one of the knock-in alleles were analysed with and without tamoxifen administration. To test whether LGR5-expressing hilum cells contribute to the rest of the OSE, we collected ovaries of *Lgr5-eGFP-IRES-CreERT2* Ai9 mice, 1 and 2 months after tamoxifen administration. The majority of OSE cells in tamoxifen but not control experiments expressed tdTomato, indicating that the hilum cells contribute to regeneration of the OSE covering the ovary (Fig. 3f and Supplementary Fig. 12k, l).

Trp53 (transformation-related protein 53) mutations and alterations of the RB1 (retinoblastoma 1) pathway occur in 96% and 67% of human high-grade serous adenocarcinomas, respectively[7,8], and result in formation of similar malignancies in genetically modified mice[27,28]. We therefore compared proliferation and immortalization efficiency of the OSE cells located at the hilum and in the remainder of the ovary after conditional inactivation of *Trp53* and *Rb1* (Supplementary Fig. 13). The hilum cells showed significantly increased proliferation and were passaged over 20 times without reaching crisis phase. In contrast, the majority of the remaining OSE cell pool became senescent by passage 6, according to their enlarged flattened shape and expression of senescence markers, such as senescence-associated β-galactosidase, p16 and p27 (Fig. 4a–c and Supplementary Fig. 14).

We evaluated various regions of the ovary carefully for the early stages of carcinogenesis induced by Cre-*loxP*-mediated recombination after a single administration of adenovirus expressing Cre (AdCre) into the ovarian bursa. Such administration results in infection of more than 80% of cells in all anatomical regions of the ovary and then formation of *Trp53*- and *Rb1*-deficient neoplastic lesions in a relatively synchronous manner (Supplementary Figs 15, 16 and ref. 27). The earliest atypical cells were detected in the hilum and adjacent areas but not in the other regions of the ovary or in the tubal epithelium in 12 out of 16 mice (75%) by 60 days after induction (Supplementary Fig. 17 and Supplementary Table 4).

To test the tumorigenic properties of *Trp53*- and *Rb1*-deficient primary OSE cells, we transplanted them intraperitoneally and subcutaneously. In intraperitoneal experiments, seven out of eight mice (87%) that were transplanted with the hilum cells developed high-grade serous adenocarcinomas characterized by extensive proliferation and expression of cytokeratin 8 (CK8), Wilms tumour protein (WT1), PAX8, oestrogen receptor 1 (ER), ephrin B1 and ALDH1 (Fig. 4d–n and Supplementary Figs 18 and 19). Notably, 5 out of 7 neoplasms (71%) metastasized to the lung (Fig. 4l–n and Supplementary Fig. 18d–f). Only 1 out of 12 mice (8%) that were transplanted with the remaining OSE cells developed carcinoma. This neoplasm expressed CK8, WT1 and ER, but contained less cells with high

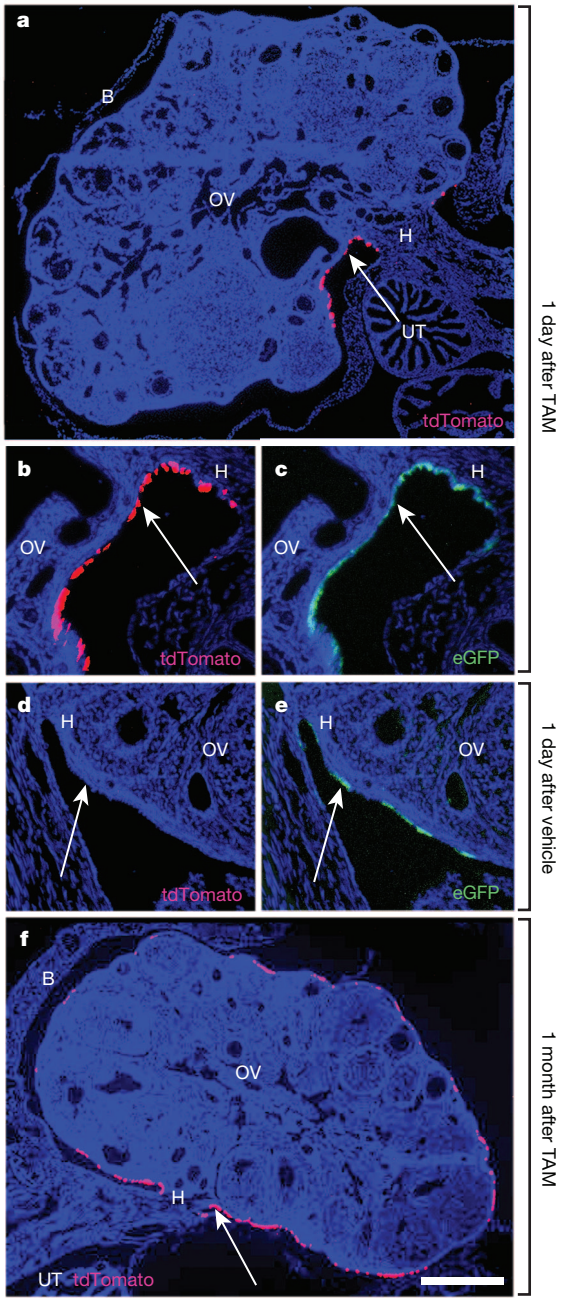


Figure 3 | Tracing the fate of LGR5$^+$ hilum cells. a–f, Detection of tdTomato (red; **a**, **b**, **d** and **f**) and eGFP (green; **c** and **e**) expression in the ovaries of *Lgr5-eGFP-IRES-CreERT2* Ai9 mice, 1 day (**a**–**e**) and 1 month (**f**) after administration of either tamoxifen (TAM; **a**–**c** and **f**) or vehicle (**d** and **e**). Counterstaining with DAPI, blue. Arrows, hilum OSE. Scale bar represents 600 μm (**a** and **f**) or 130 μm (**b**–**e**).

ALDH1 expression and only a few PAX8-positive cells, and it was not metastatic (Supplementary Figs 19 and 20). Similar results were also observed in subcutaneous experiments (Supplementary Figs 19–21 and Supplementary Table 5). Thus, the hilum cells have increased transformation potential after inactivation of *Trp53* and *Rb1* and may be the cell of origin of EOC.

In summary, we show that the OSE at junction areas contains a novel stem cell niche that is responsible for OSE regeneration and is prone to malignant transformation. These findings provide experimental rationale for targeted detection and characterization of pre-neoplastic and early neoplastic lesions in the areas of transition between OSE, mesothelium and tubal epithelium in humans. Previously, presence of

©2013 Macmillan Publishers Limited. All rights reserved

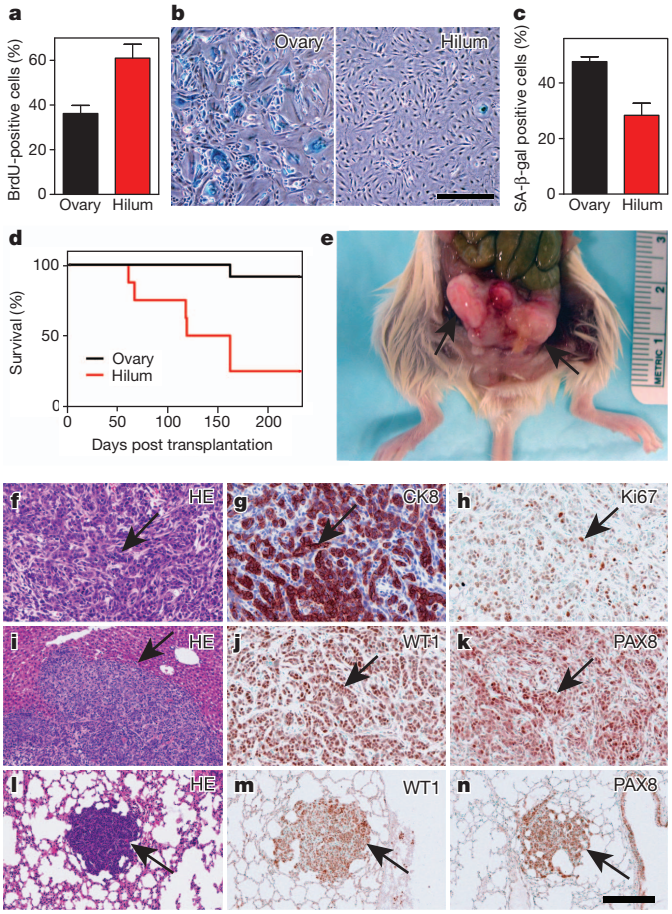


Figure 4 | Hilum cells show preferential transformation after conditional inactivation of *Trp53* and *Rb1*. Characterization of primary OSE cells isolated from the hilum and the remainder of the ovary, and evaluated at passage 6 after Cre-*loxP*-mediated inactivation of *Trp53* and *Rb1*. **a**, Quantification of BrdU-positive OSE cells ($n = 3$, $P < 0.0001$, two-tailed unpaired *t*-test). **b**, **c**, Detection (**b**) and quantitative analysis (**c**) of senescence-associated β-galactosidase (SA-β-gal; blue) in the ovary and hilum cells ($n = 3$, $P = 0.002$, two-tailed unpaired *t*-test), using phase contrast microscopy. Scale bar in **b**, 500 μm. **d**, Survival of mice that were intraperitoneally transplanted with primary ovary ($n = 12$) and hilum ($n = 8$) OSE cells deficient for *Trp53* and *Rb1* (log-rank $P = 0.0007$). **e**, Neoplastic masses (arrows) from hilum cells transplanted into the mouse abdominal cavity. **f**–**n**, Hilum-derived neoplastic cells have marked nuclear atypia (**f**), express cytokeratin 8 (CK8, **g**), are highly proliferative (**h**), grow in solid, nested and gland-like patterns (**f**–**h**, **j** and **k**, arrows), invade liver (**i**, arrow), metastasize to the lung (**l**–**n**, arrows) and express Wilms tumour protein (WT1; **j** and **m**) and PAX8 (**k** and **n**). Haematoxylin and eosin (HE) and immunostaining, Elite ABC method. Counterstaining with haematoxylin (CK8) and methyl green (Ki67, WT1 and PAX8). Scale bar represents 100 μm (**f**–**h**, **j** and **k**) or 200 μm (**i** and **l**–**n**). All error bars denote s.d.

adult stem cells in the areas of a junction between two types of epithelia was demonstrated definitively for the limbus region, a narrow transitional zone between the cornea and bulbar conjunctiva[29], and for the gastroesophageal junction[30]. However, it remains unclear whether such junction-associated stem cell niches are predisposed to cancer. Our findings support this possibility and suggest that similar junction areas in other organs, such as the uterine cervix and the anus, may also contain cancer-prone stem cell niches, thereby explaining susceptibility of such regions to malignant transformation. Given the anatomically well-defined location of the hilum in the mouse ovary, this region may represent an attractive model for further studies aimed at understanding why stem cell niches reside in the junction areas and how aberrations in the molecular and cellular mechanisms governing epithelium regeneration may contribute to ovarian carcinoma pathogenesis.

METHODS SUMMARY

Primary antibodies used for immunohistochemistry are listed in Supplementary Table 6. Cell culture, molecular, histology and pathology techniques, as well as carcinogenesis experiments, were carried out as described previously[27] and in the Methods section.

Full Methods and any associated references are available in the online version of the paper.

Received 19 January 2012; accepted 1 February 2013.
Published online 6 March 2013.

1. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2012. *CA Cancer J. Clin.* **62,** 10–29 (2012).
2. Lengyel, E. Ovarian cancer development and metastasis. *Am. J. Pathol.* **177,** 1053–1064 (2010).
3. Bowtell, D. D. The genesis and evolution of high-grade serous ovarian cancer. *Nature Rev. Cancer* **10,** 803–808 (2010).
4. Visvader, J. E. Cells of origin in cancer. *Nature* **469,** 314–322 (2011).
5. Auersperg, N. The origin of ovarian carcinomas: a unifying hypothesis. *Int. J. Gynecol. Pathol.* **30,** 12–21 (2011).
6. Medema, J. P. & Vermeulen, L. Microenvironmental regulation of stem cells in intestinal homeostasis and cancer. *Nature* **474,** 318–326 (2011).
7. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615 (2011).
8. Ahmed, A. A. *et al.* Driver mutations in *TP53* are ubiquitous in high grade serous carcinoma of the ovary. *J. Pathol.* **221,** 49–56 (2010).
9. Kurman, R. J. & Shih Ie, M. The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am. J. Surg. Pathol.* **34,** 433–443 (2010).
10. Dubeau, L. The cell of origin of ovarian epithelial tumours. *Lancet Oncol.* **9,** 1191–1197 (2008).
11. Auersperg, N., Woo, M. M. & Gilks, C. B. The origin of ovarian carcinomas: a developmental view. *Gynecol. Oncol.* **110,** 452–454 (2008).
12. Seidman, J. D., Yemelyanova, A., Zaino, R. J. & Kurman, R. J. The fallopian tube-peritoneal junction: a potential site of carcinogenesis. *Int. J. Gynecol. Pathol.* **30,** 4–11 (2011).
13. Szotek, P. P. *et al.* Normal ovarian surface epithelial label-retaining cells exhibit stem/progenitor cell characteristics. *Proc. Natl Acad. Sci. USA* **105,** 12469–12473 (2008).
14. Blanpain, C., Horsley, V. & Fuchs, E. Epithelial stem cells: turning over new leaves. *Cell* **128,** 445–458 (2007).
15. Simons, B. D. & Clevers, H. Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell* **145,** 851–862 (2011).
16. Ginestier, C. *et al.* ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1,** 555–567 (2007).
17. Storms, R. W. *et al.* Isolation of primitive human hematopoietic progenitors on the basis of aldehyde dehydrogenase activity. *Proc. Natl Acad. Sci. USA* **96,** 9118–9123 (1999).
18. Corti, S. *et al.* Identification of a primitive brain-derived neural stem cell population based on aldehyde dehydrogenase activity. *Stem Cells* **24,** 975–985 (2006).
19. Burger, P. E. *et al.* High aldehyde dehydrogenase activity: a novel functional marker of murine prostate stem/progenitor cells. *Stem Cells* **27,** 2220–2228 (2009).
20. Deng, S. *et al.* Distinct expression levels and patterns of stem cell marker, aldehyde dehydrogenase isoform 1 (ALDH1), in human epithelial cancers. *PLoS ONE* **5,** e10277 (2010).
21. Bullough, W. S. Oogenesis and its relation to the oestrous cycle in the adult mouse. *J. Endocrinol.* **3,** 141–149 (1942).
22. Ji, Q. *et al.* MicroRNA miR-34 inhibits human pancreatic cancer tumor-initiating cells. *PLoS ONE* **4,** e6816 (2009).
23. Liu, C. *et al.* The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44. *Nature Med.* **17,** 211–215 (2011).
24. Choi, Y. J. *et al.* miR-34 miRNAs provide a barrier for somatic cell reprogramming. *Nature Cell Biol.* **13,** 1353–1360 (2011).
25. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449,** 1003–1007 (2007).
26. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13,** 133–140 (2010).
27. Flesken-Nikitin, A., Choi, K. C., Eng, J. P., Shmidt, E. N. & Nikitin, A. Y. Induction of carcinogenesis by concurrent inactivation of *p53* and *Rb1* in the mouse ovarian surface epithelium. *Cancer Res.* **63,** 3459–3463 (2003).
28. Szabova, L. *et al.* Perturbation of Rb, p53, and Brca1 or Brca2 cooperate in inducing metastatic serous epithelial ovarian cancer. *Cancer Res.* **72,** 4141–4153 (2012).
29. Pellegrini, G. *et al.* Location and clonal analysis of stem cells and their differentiated progeny in the human ocular surface. *J. Cell Biol.* **145,** 769–782 (1999).
30. Barker, N. *et al.* Lgr5^{+ve} stem cells drive self-renewal in the stomach and build long-lived gastric units *in vitro*. *Cell Stem Cell* **6,** 25–36 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank T. Tumbar for critical reading of this manuscript, J. Choi for help with immunostainings, L. Sayam (NYSTEM supported FACS Core) for help with FACS experiments and F. M. Vermeylen (Cornell Statistical Consulting Unit) for help with statistical analysis. This work was supported by grants

©2013 Macmillan Publishers Limited. All rights reserved

from the US National Institutes of Health (NIH) and National Cancer Institute (NCI) (CA096823 and CA112354), NYSTEM (C023050 and N11G-160) and the Marsha Rivkin Center for Ovarian Cancer Research to A.Yu.N.; the NIH National Institute of Mental Health (MH092928), NIH National Institute on Aging (AG040209), NYSTEM (C024323) and Russian Ministry of Education and Science to G.E.; and postdoctoral fellowships to A.F.-N. (NIH NICHD T32HD052471) and C-.I.H. (Cornell Comparative Cancer Biology Training Program).

Author Contributions A.F.-N. and A.Yu.N. designed the study, interpreted data and wrote the manuscript. A.F.-N., C.-I.H., C.-Y.C., T.V.M. and G.E. performed experiments and analysed data. All authors discussed the results and commented on the manuscript.

Author Information Microarray data discussed in this publication have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus and are accessible through GEO Series accession number GSE43897. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.Yu.N. (an58@cornell.edu).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Experimental animals. FVB/NCr (FVB/N) and Swiss Webster (Cr:SW) mice were obtained from NCI-Frederick Animal Production Program (Charles River Laboratories) or bred in-house. C57BL/6J (B6) mice and B6 × 129 crosses were bred in-house. The C57BL/6-Tg(CAG-eGFP)1Osb/J (*β-actin-eGFP*), B6.Cg-Tg(CAG-DsRed*MST)1Nagy/J (*β-actin-DsRed*), *Lgr5*$^{tm1(cre/ERT2)Cle}$/J (*Lgr5-eGFP-IRES-CreERT2*), *Gt(ROSA)26Sor*$^{tm9(CAG-tdTomato)Hze}$/J (Ai9) and *NOD.Cg-Prkdc*^{scid}*Il2rg*^{tm1Wjl}/*SzJ* mice were obtained from The Jackson Laboratory. *Trp53*$^{loxP/loxP}$ and *Rb1*$^{loxP/loxP}$ mice (*Trp53* and *Rb1* genes carrying floxed alleles) were a gift from A. Berns. All mice were maintained identically, following recommendations of the Institutional Laboratory Animal Use and Care Committee.

Isolation of primary OSE cells. Individual ovaries were dissected from 6- to-8-week-old virgin FVB/N or wild-type mice, transferred to 100 µl digestion buffer (4 mg Collagenase-Dispase (Roche) per ml DMEM/F12 (Ham's) medium (Mediatech) supplemented with 30 mg ml^{-1} bovine albumin (Sigma) and 1 µl DNase I (1 mg ml^{-1}, Sigma)), incubated for 55 min at 37 °C in a 5% CO_2 incubator. After the incubation, ovaries were removed and 5 ml complete OSE stem cell medium (OSE SCM) was added. OSE SCM consists of 5% fetal bovine serum (FBS; Sigma), 4 mM L-glutamine (Mediatech), 1 mM sodium pyruvate (Mediatech), 10 ng ml^{-1} epidermal growth factor (Sigma), 500 ng ml^{-1} hydrocortisone (Sigma), 5 µg ml^{-1} insulin (Sigma), 5 µg ml^{-1} transferrin (Sigma), 5 ng ml^{-1} sodium selenite (Sigma), 0.1 mM MEM non-essential amino acids (Mediatech), 10^{-4} M β-mercaptoethanol (Sigma), 10^3 units per ml leukaemia inhibitory factor (Millipore) in DMEM/F12 (Ham's) medium (Mediatech). Cells were collected by centrifugation, and 2×10^4 cells were seeded onto one gelatinized 24-well plate and incubated in OSE SCM at 5% CO_2 for 24 h to eliminate accidental non-OSE cells[27,31,32]. The purity of isolated cells was confirmed by RT–PCR (PCR with reverse transcription) detection of CK8 expression.

ALDEFLUOR assay and fluorescence-activated cell sorting. For detection of ALDH enzymatic activity, primary OSE cells (4×10^6) were placed in ALDEFLUOR buffer and processed for staining with the ALDEFLUOR kit (Aldagen) according to the manufacturer's protocol. Cell sorting and data analysis were performed on a FACS Aria II sorter equipped with the FACS DiVa software (BD Bioscience). Unstained cells and cells treated with ALDH inhibitor (4-(diethylamino)benzaldehyde[33] at a tenfold molar excess) served as controls. Dead cells were excluded with propidium iodide. The brightest 5–8% of ALDH$^+$ cells were identified and gated electronically based on their characteristic light-scatter properties on the fluorescein isothiocyanate (FITC)-channel emission pattern after excitation with a 13–20 mW, 488-nm ellipse-shaped laser (elliptical) BD FACSAria II. The ALDH fluorescence emissions were captured simultaneously through a 515/20-nm band-pass and 505-nm long-pass filter. ALDH$^+$ and ALDH$^-$ OSE cells were collected in 0.5 ml OSE SCM and subjected to OSE-sphere rim assays or RNA isolation, as described below.

OSE-sphere-formation rim assay. Cell populations (5×10^5 cells per assay) were collected in 0.5 ml OSE SCM and centrifuged, and pellets were suspended in reduced growth-factor basement-membrane matrix (Geltrex, Invitrogen) and OSE SCM (1:1 ratio) in a total volume of 120 µl. A modified protocol of Lawson *et al.*[34] was followed; cells were placed around the rim of a well of a 12-well plate and allowed to solidify for 25 min at 37 °C in a 5% CO_2 incubator before adding 1.5 ml OSE SCM. Spheres were grown for 7 to 12 days. For passaging of spheres, media were aspirated and Geltrex was digested by incubation in 750 µl digestion buffer for 1 h, at 37 °C. During incubation, spheres were suspended 2 to 3 times manually by pipetting through a blue 1-ml pipette tip. Digested cultures were pelleted and incubated in 0.3 ml 0.25% trypsin/EDTA for 10 min at 37 °C, cells were suspended and the enzyme reaction was stopped by adding 4 ml OSE SCM. Cells were collected and counted using a haemocytometer, and then cultured at different densities.

Confocal imaging of OSE-sphere suspension culture. Primary OSE cells from *β-actin-eGFP* mice were grown for 4 or 10 days in low-attachment 24-wells (Corning) at 2×10^4 cells per ml OSE stem cell conditioned medium, which was composed of the following: (DMEM/F12 (Ham's) medium (Mediatech) containing 2.4% methylcellulose (Sigma), 0.4% BSA (Sigma), 4 mM L-glutamine (Mediatech), 1 mM sodium pyruvate (Mediatech), 10 ng ml^{-1} epidermal growth factor (Sigma), 500 ng ml^{-1} hydrocortisone (Sigma), 5 µg ml^{-1} insulin (Sigma), 5 µg ml^{-1} transferrin (Sigma), 5 ng ml^{-1} sodium selenite (Sigma), 10 ng ml^{-1} fibroblast growth factor basic (Sigma), 0.1 mM MEM non-essential amino acids (Mediatech), 10^3 units per ml leukaemia inhibitory factor (Millipore), 10^{-4} M β-mercaptoethanol (Sigma). Using a drawn-out glass capillary under a dissection microscope, spheres were washed twice for 3 min in a drop of PBS at room temperature (20–22 °C), fixed in 4% paraformaldehyde for 5 min, washed twice with PBS, once with ddH_2O, stained with DAPI (Sigma, 1 mg ml^{-1}) for 5 min, transferred to a glass slide, mounted in Fluorogel (Electron Microscopy Sciences) and covered by a glass coverslip. Confocal pictures were acquired at room temperature using a Zeiss LSM 510 Meta confocal microscope coupled with an Axiovert 200 inverted microscope (Carl Zeiss). Optical sections, 512 × 512 pixels, were collected sequentially for each fluorochrome. The data sets acquired were merged and displayed with the ZEN software (Zeiss).

Monoclonality assays. Two approaches were used. First, similarly to earlier studies in other cell types[34–36], single-cell suspensions of primary OSE cells from *β-actin-eGFP* or *β-actin-DsRed* mice were mixed at various ratios, grown in OSE-sphere-forming rim assays for 7–12 days without subculture in stem cell conditioned medium, and then the numbers of monochromatic and dichromatic spheres were counted. Second, the development of spheres from individual cells in rim assays was monitored by imaging the same fields of view at different time points. All images were collected using a Nikon Eclipse TS100 microscope coupled with an Epi-Fluorescence Attachment (Nikon), equipped with a DS-L2 camera control unit and cooled camera head DS-5Mc (Nikon).

Primary culture of regions of the hilum and ovary. Individual hilum and anterior ovary regions were isolated under dissection microscope from adult virgin FVB/N mice, minced in a drop of PBS with 25G needles and transferred to 100 µl of digestion buffer. Samples were then incubated as described for the isolation of primary OSE cells. After 24 h, adherent cultures of cells from the hilum and anterior ovary region were collected and seeded out for the cell proliferation or OSE-sphere rim assays.

Gene-expression arrays. Total RNA was isolated using a mirVana miRNA isolation kit (Ambion). An Ovation Pico WTA System V2 (NuGen) was used to amplify complementary DNA, and this was then labelled with biotin using Encore Biotin Module (NuGen). Labelled DNA was hybridized onto an Affymetrix Mouse Genome 430 2.0 array (Affymetrix). All microarray analyses were carried out as described previously[31,32,37,38]. In briefly, microarray data were analysed with GeneSifter software (Geospiza). To identify genes that were significantly altered in ALDH-positive populations, we carried out paired two-group analysis using Significance Analysis of Microarrays (SAM) software (http://www-stat.stanford.edu/~tibs/SAM/) and visualizations using Treeview software (http://rana.lbl.gov).

***Trp53* and *Rb1* inactivation in primary OSE cells.** Individual samples of hilum regions, and all ovary regions except the hilum, were isolated under a dissection microscope from adult *Trp53*$^{loxP/loxP}$*Rb1*$^{loxP/loxP}$ mice, and then incubated as described for the isolation of primary OSE cells. Cells were passaged one or two times and treated with recombinant adenoviruses essentially as described previously[27,31]. No changes in composition of ALDH1$^+$- and ALDH$^-$-cell subpopulations were observed between cells tested 24 h after initial adherent culture and cells tested immediately before exposure to adenovirus. At passage 6 after *Trp53* and *Rb1* inactivation, proliferation- and senescence-detection assays were carried out as described below.

Cell-proliferation assays. Primary cells from the hilum and ovary regions, passage 0, were collected, reseeded in triplicate at 1×10^4 cells per 3.5-cm dish in OSE SCM, grown for 3 days, then stained with Giemsa solution (Sigma), air-dried and analysed in a total area of 4.92 cm^2. For BrdU staining, cells were incubated with 3 µg ml^{-1} BrdU for 2 h at 37 °C and fixed with 4% paraformaldehyde. DNA was denatured using 4N HCl with 0.5% Triton X-100 (Sigma) at room temperature for 10 min, and cells were washed in Tris-buffered saline with 0.5% Triton X-100. Primary antibody for BrdU (Supplementary Table 6) was applied overnight at 4 °C, and cells were then incubated with secondary biotinylated antibody (1 h at room temperature) and a modified avidin–biotin–peroxidase technique was used as described previously[39].

Senescence-detection assays. For detection of senescence-associated β-galactosidase, OSE cells were fixed with 2% formaldehyde/0.2% glutaraldehyde for 5 min, rinsed several times with PBS, and incubated in senescence-associated β-galactosidase staining solution (4.2 mM citric acid, 12.5 mM sodium phosphate, 158 mM sodium chloride, 0.21 mM magnesium chloride, 2.21 mg ml^{-1} potassium ferrocyanide, 1.68 mg ml^{-1} potassium ferricyanid (all Sigma), 1 mg ml^{-1} 5-bromo-4-chloro-3-indolyl β-D-galoctosidase (X-Gal, Invitrogen), pH 6.0) for 24 h at 37 °C according to a published protocol[40]. For immunofluorescent detection of p16 and p27, cells were fixed with 4% paraformaldehyde and incubated with primary antibodies (Supplementary Table 6) overnight at 4 °C. Secondary antibody (Alexa Fluor 488 conjugate, Invitrogen) was applied for 1 h at room temperature. DAPI was used for nuclear counterstaining.

BrdU pulse-chase experiments. Six- to seven-week-old virgin FVB/N mice were injected daily with 250 µl of BrdU (Sigma; 1 mg ml^{-1}) intraperitoneally for 10 days (pulse) and euthanized immediately after the pulse, or 1, 2 or 3 months after pulse, in 3 independent experiments. For LCR-proliferation analysis during the oestrus cycle, the stage of the oestrus cycle was checked by vaginal cytology and mice were euthanized either during pro-oestrus (before ovulation) or oestrus (after ovulation), after 2 or 3 months of chase.

©2013 Macmillan Publishers Limited. All rights reserved

Cell-lineage tracing. To deliver the tamoxifen pulse, 6- to 7-week-old virgin *Lgr5-eGFP-IRES-CreERT2* Ai9 and control mice were injected intraperitoneally with a single dose of tamoxifen (Sigma, 0.2 mg g^{-1} body weight, 25 mg ml^{-1} in corn oil) or three doses every other day for 6 days. Ovaries of tamoxifen-pulsed mice were collected 1 and 3 days after the pulse, and at monthly intervals for 2 months. Collected ovaries were fixed with 4% paraformaldehyde for 2 h at room temperature, washed twice for 3 min with PBS, kept overnight at 4 °C, embedded in Tissue-Tek OCT compound (Sakura Fineted) and frozen in liquid nitrogen. Frozen sections were counterstained with DAPI (Sigma, 1 mg ml^{-1}) and images were taken with an Axioskop 2 fluorescence microscope (Zeiss) equipped with a charge-coupled device (CCD) camera.

Intrabursal administration of adenovirus and carcinogenesis induction. At oestrus, adult *Trp53*$^{loxP/loxP}$*Rb1*$^{loxP/loxP}$ mice received a single trans-infundibular intrabursal injection of recombinant adenovirus AdCre-eGFP, AdCre, AdeGFP or AdBlank (Gene Transfer Vector Core) as described previously[27]. The original stock solution (stored at −80 °C) with a titre of 7×10^{10} plaque-forming units per ml was diluted at a ratio of 1:20 with PBS at room temperature immediately before the injection.

Laser microdissection and PCR genotyping. For microdissection of frozen sections, 5-μm-thick frozen sections of the mouse ovaries were placed on slides for membrane-based laser microdissection (Leica Microsystems) and evaluated under the Leica DM LA microscope coupled with a Fluorescence illuminator LRF 4/22. eGFP-expressing OSE from the hilum and different ovarian regions were microdissected using the Leica AS LMD laser-microdissection system (Leica) equipped with a VSL-337ND-S nitrogen laser. Phase-contrast and fluorescent images were acquired before and after the microdissection. Manual microdissection of haematoxylin-and-eosin-stained sections was carried out essentially as described preciously[27,39,41]. Tissues were collected into caps of 0.6-ml centrifuge tubes filled with lysis buffer and used for subsequent PCR amplification prepared as described previously[27,37,39,42,43].

Tumorigenicity assays. Primary OSE cells deficient for *Trp53* and *Rb1* were derived from primary hilum, or all ovary regions except the hilum, and were suspended in PBS and injected intraperitoneally (8×10^6 cells in 500 μl) or subcutaneously ($5 \times 10^4 - 5 \times 10^6$ cells in 250 μl) into 5-week-old *NOD.Cg-Prkdc*^{scid}*Il2rg*^{tm1Wjl}/*SzJ4* female mice. As soon as signs of sickness were seen (for example, abdominal distention, intra-abdominal masses or subcutaneous masses over 1 cm in diameter or moribund behaviour), mice were euthanized and subjected to necropsy.

Morphological evaluation. Owing to the coiled structure of the mouse uterine tubes[44], 4-μm-thick paraffin serial sections within 100 μm of the junction between the OSE and tubal epithelium were used for all studies of the hilum region. Serial sections were also evaluated to exclude potential misinterpretation of tangential sections in all relevant quantitative assays. All mice in carcinogenesis experiments were subjected to gross pathology evaluation during necropsy. Particular attention was paid to potential sites of ovarian carcinoma spreading, such as the omentum, regional lymph nodes, liver, lung and mesentery. In addition to the ovary, pathologically altered organs, as well as representative specimens of the brain, lung, liver, kidney, spleen, pancreas and intestine, omentum and intra-abdominal lymph nodes, were fixed in 4% PBS buffered paraformaldehyde. They were then evaluated by microscopic analysis of paraffin sections stained with haematoxylin and eosin, and subjected to necessary immunostainings. All early atypical lesions were diagnosed based on their morphology, staining for Ki67 and detection of deleted (floxed-out) *Trp53* and *Rb1* as described earlier[27]. The locations of all lesions were determined by three-dimensional reconstruction of 4-μm-thick serial sections as described previously[39].

Immunohistochemistry and quantitative image analysis. Information regarding all primary antibodies is provided in Supplementary Table 6. Immunohistochemical analysis of paraffin sections of paraformaldehyde-fixed tissue was carried out using a modified ABC technique[37–39]. For immunofluorescence analysis, de-paraffinized or frozen sections and cells were fixed in 4% paraformaldehyde. Ten-minute treatment with 4N HCl and 10 min boiling in 10 mM citric buffer was used to retrieve antigen for detection of BrdU. Staining of theca cells[45] served as a positive internal control for ALDH1 detection. Sections with immunoperoxidase stainings were scanned by ScanScope (Aperio Technologies) with a ×40 objective followed by lossless compression. Sections with immunofluorescent stainings were analysed under an Axioskop 2 (Zeiss) fluorescence microscope equipped with a CCD camera. Quantitative analysis of immunohistochemistry and immunofluorescence experiments were carried out using the ImageJ software (NIH) as described previously[37–39].

Quantitative real-time PCR. All procedures were carried out as described previously[32], with at least three separately prepared RNA samples for each of three independent experiments. A list of the genes tested is included in Supplementary Table 3. The following primers for detection of microRNAs were obtained from Quanta Biosciences: U6 small nuclear RNA (catalogue no. HSRNU6); miR-34a (catalogue no. HSMIR-0034A); miR-34b (catalogue no. MMIR-0034B-5P); miR-34c (catalogue no. HSMIR-0034C-3P); miR-376b (catalogue no. MMIR-0376B).

Statistical analyses. All statistical analyses in this study were carried out using InStat 3.10 and Prism 5.04 software (GraphPad) as described previously[27,37].

31. Corney, D. C., Flesken-Nikitin, A., Godwin, A. K., Wang, W. & Nikitin, A. Y. *MicroRNA-34b* and *microRNA-34c* are targets of p53 and cooperate in control of cell proliferation and adhesion-independent growth. *Cancer Res.* **67,** 8433–8438 (2007).
32. Hwang, C. I. *et al.* Wild-type p53 controls cell motility and invasion by dual regulation of MET expression. *Proc. Natl Acad. Sci. USA* **108,** 14240–14245 (2011).
33. Russo, J. E., Hauguitz, D. & Hilton, J. Inhibition of mouse cytosolic aldehyde dehydrogenase by 4-(diethylamino)benzaldehyde. *Biochem. Pharmacol.* **37,** 1639–1642 (1988).
34. Lawson, D. A., Xin, L., Lukacs, R. U., Cheng, D. & Witte, O. N. Isolation and functional characterization of murine prostate stem cells. *Proc. Natl Acad. Sci. USA* **104,** 181–186 (2007).
35. Xin, L., Lukacs, R. U., Lawson, D. A., Cheng, D. & Witte, O. N. Self-renewal and multilineage differentiation in vitro from murine prostate stem cells. *Stem Cells* **25,** 2760–2769 (2007).
36. Singec, I. *et al.* Defining the actual sensitivity and specificity of the neurosphere assay in stem cell biology. *Nature Methods* **3,** 801–806 (2006).
37. Zhou, Z. *et al.* Synergy of p53 and Rb deficiency in a conditional mouse model for metastatic prostate cancer. *Cancer Res.* **66,** 7889–7898 (2006).
38. Choi, J., Curtis, S. J., Roy, D. M., Flesken-Nikitin, A. & Nikitin, A. Y. Local mesenchymal stem/progenitor cells are a preferential target for initiation of adult soft tissue sarcomas associated with p53 and Rb deficiency. *Am. J. Pathol.* **177,** 2645–2658 (2010).
39. Nikitin, A. Y. & Lee, W. H. Early loss of the retinoblastoma gene is associated with impaired growth inhibitory innervation during melanotroph carcinogenesis in $Rb^{+/-}$ mice. *Genes Dev.* **10,** 1870–1879 (1996).
40. Debacq-Chainiaux, F., Erusalimsky, J. D., Campisi, J. & Toussaint, O. Protocols to detect senescence-associated beta-galactosidase (SA-βgal) activity, a biomarker of senescent cells in culture and *in vivo*. *Nature Protocols* **4,** 1798–1806 (2009).
41. Matoso, A., Zhou, Z., Hayama, R., Flesken-Nikitin, A. & Nikitin, A. Y. Cell lineage-specific interactions between Men1 and Rb in neuroendocrine neoplasia. *Carcinogenesis* **29,** 620–628 (2008).
42. Zhou, Z., Flesken-Nikitin, A. & Nikitin, A. Y. Prostate cancer associated with p53 and Rb deficiency arises from the stem/progenitor cell-enriched proximal region of prostatic ducts. *Cancer Res.* **67,** 5683–5690 (2007).
43. Zhou, Z. *et al.* Suppression of melanotroph carcinogenesis leads to accelerated progression of pituitary anterior lobe tumors and medullary thyroid carcinomas in $Rb^{+/-}$ mice. *Cancer Res.* **65,** 787–796 (2005).
44. Rugh, R. In *The Mouse: Its Reproduction and Development* 7–43 (Oxford Univ. Press, 1990).
45. Vermot, J., Fraulob, V., Dolle, P. & Niederreither, K. Expression of enzymes synthesizing (aldehyde dehydrogenase 1 and reinaldehyde dehydrogenase 2) and metabolizaing (Cyp26) retinoic acid in the mouse female reproductive system. *Endocrinology* **141,** 3638–3645 (2000).

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature11912

Naturally occurring allele diversity allows potato cultivation in northern latitudes

Bjorn Kloosterman[1]†, José A. Abelenda[2]*, María del Mar Carretero Gomez[1]*, Marian Oortwijn[1], Jan M. de Boer[1], Krissana Kowitwanich[1], Beatrix M. Horvath[3], Herman J. van Eck[1,4], Cezary Smaczniak[5], Salomé Prat[2], Richard G. F. Visser[1,4] & Christian W. B. Bachem[1]

Potato (*Solanum tuberosum* L.) originates from the Andes and evolved short-day-dependent tuber formation as a vegetative propagation strategy. Here we describe the identification of a central regulator underlying a major-effect quantitative trait locus for plant maturity and initiation of tuber development. We show that this gene belongs to the family of DOF (DNA-binding with one finger) transcription factors[1] and regulates tuberization and plant life cycle length, by acting as a mediator between the circadian clock and the StSP6A mobile tuberization signal[2]. We also show that natural allelic variants evade post-translational light regulation, allowing cultivation outside the geographical centre of origin of potato. Potato is a member of the Solanaceae family and is one of the world's most important food crops. This annual plant originates from the Andean regions of South America[3]. Potato develops tubers from underground stems called stolons. Its equatorial origin makes potato essentially short-day dependent for tuberization and potato will not make tubers in the long-day conditions of spring and summer in the northern latitudes. When introduced in temperate zones, wild material will form tubers in the course of the autumnal shortening of day-length. Thus, one of the first selected traits in potato leading to a European potato type[4] is likely to have been long-day acclimation for tuberization. Potato breeders can exploit the naturally occurring variation in tuberization onset and life cycle length, allowing varietal breeding for different latitudes, harvest times and markets.

The timing of reproductive development is determined by environmental cues[5]. In *Arabidopsis thaliana*, flowering is initiated by the expression and transport of a peptide hormone or 'florigen' called FLOWERING LOCUS T (FT)[6]. Orthologues of *FT* have been shown to regulate flowering in diverse plant species, including the short-day plant rice (*HD3A*)[7], flowering day-length neutral tomato and potato (*SP3D*)[2,8] and the biannual sugar beet (*BvFT1* and *BvFT2*)[9]. An additional potato homologue of *FT* (*StSP6A*) has recently been demonstrated to be the mobile 'tuberigen' signal that initiates tuberization and this is regulated separately from the flowering pathway[2]. In *A. thaliana*, *FT* expression is controlled by the diurnally regulated transcription factor CONSTANS (CO)[10]. Activity of the CO protein is post-transcriptionally regulated by light through an interaction with the red and blue light receptors, phytochrome A, B, and cryptochrome 2, respectively (ref. 11). CO acts downstream of a signalling cascade that involves interaction of the clock gene GIGANTEA (GI) (ref. 12) and a number of other blue-light absorbing proteins such as FLAVIN-BINDING KELCH REPEAT F-BOX PROTEIN 1 (FKF1), LOV KELCH REPEAT PROTEIN 2 (LKP2) and ZEITLUPPE (ZTL) that regulate downstream effectors, including members of the CYCLING DOF FACTOR (CDF) family[13]. Binding of FKF1 and GI to the CDF protein targets it for degradation by the proteasome[14]. In *A. thaliana*, CDF1 binds the *CO* and *FT* promoters, leading to the inhibition of flowering[14,15]. Potato homologues of *CONSTANS* (*StCO*) were also shown to be involved in day-length tuberization control in combination with the red light receptor phytochrome B (*StPHYB*)[16,17].

The potato plant maturity phenotype is observed as a major-effect quantitative trait locus (QTL) that maps to chromosome 5 (ref. 18). Breeders score this phenotype as a measure for several important subtraits including the onset of leaf senescence and total length of the plant life cycle, as well as the associated time point of tuberization and ultimately potato yield.

We used two diploid potato populations, derived from wild and domesticated genotypes (generated from the parental crosses C × E and SH × RH; see detailed description in Supplementary Information) to map the plant maturity QTL (Fig. 1a). To define the QTL region better, an additional set of nearly 2,000 individual siblings from the C × E and SH × RH crosses were generated and screened with existing genetic markers to identify recombination locations in this genetic region. Subsequent phenotyping and fine mapping (see Methods) allowed us to narrow down the locus responsible for this trait to a region of around 110 kilobases (kb) (Fig. 1a, Supplementary Information and Supplementary Fig. 1b). As the diploid RH genome has a good coverage of sequenced and annotated bacterial artificial chromosomes (BACs) for both homologous chromosomes in this region[19], we screened for possible candidate genes and identified a *CDF* homologue (PGSC003DMG400018408) on chromosome 5, hereafter referred to as *StCDF1* (*Solanum tuberosum CDF* gene 1), as a candidate gene for the plant maturity phenotype based on its regulatory links with the *StCO* and *StSP6A* genes that have an established role in tuberization[20]. Other genes in this region showed no phenotypic effects on tuberization when overexpressed in transgenic plants (Supplementary Information).

Sequence analysis of the *StCDF1* gene in the very late-maturing C × E descendant CE3027 showed homozygosity for an *StCDF1* allele denoted *StCDF1.1* (Supplementary Fig. 2a–c), whereas the very early maturing descendant CE3130 has two different truncated alleles (denoted *StCDF1.2* and *StCDF1.3*; Supplementary Fig. 2a–c). In *StCDF1.2,* an insertion results in a frame-shift introducing a premature stop codon (7-base-pair (bp) insertion; Supplementary Information and Supplementary Fig. 2a), whereas in *StCDF1.3* there is a larger insertion (865 bp; Supplementary Information) that leads to a fusion protein with 22 additional amino acids, replacing the 52 carboxy-terminal residues found in the StCDF1.1 protein (Supplementary Fig. 2a).

The early maturing/tuberizing *StCDF1.2* gene variant, under control of the cauliflower mosaic virus *35S* promoter (CaMV *35S*), was introduced into the short-day-dependent *S. tuberosum* group *andigenum* (referred to here as *S. andigenum*) genotype[21] and into the very late maturing CE3027 descendant. Under long-day conditions transformed

[1]Laboratory of Plant Breeding, Department of Plant Sciences, Wageningen-UR, PO Box 386, 6700 AJ Wageningen, The Netherlands. [2]Departamento de Genética Molecular de Plantas, Centro Nacional de Biotecnología-CSIC, Darwin 3, 28049 Madrid, Spain. [3]Molecular Genetics, Utrecht University, PO Box 80125, 3508 TC Utrecht, The Netherlands. [4]Centre for BioSystems Genomics, PO Box 98, 6700 AB Wageningen, The Netherlands. [5]Laboratory of Molecular Biology, Department of Plant Sciences, Wageningen-UR, PO Box 386, 6700 AJ Wageningen, The Netherlands. †Present address: Bjorn Kloosterman, Keygene NV, PO Box 216, 6700 AE Wageningen, The Netherlands.
*These authors contributed equally to this work.

©2013 Macmillan Publishers Limited. All rights reserved

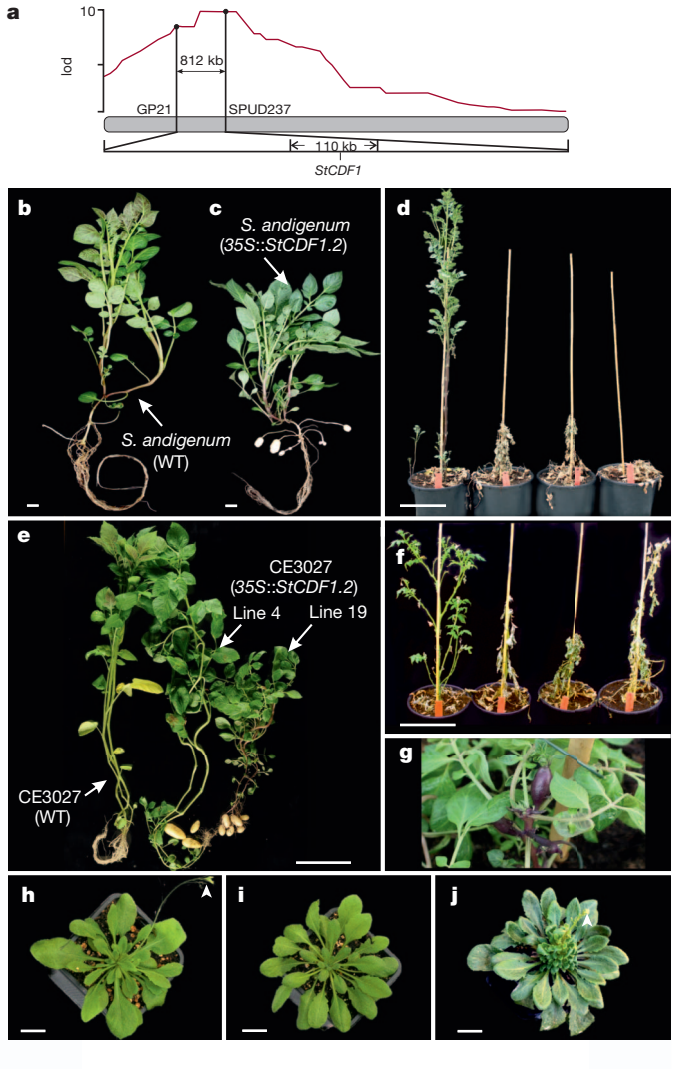


Figure 1 | Overexpression of *StCDF1.2*, localized on chromosome 5, results in earlier potato tuber formation, and delays flowering in *Arabidopsis*. **a**, Localization of the plant maturity QTL on chromosome 5. The red graph trace represents the logarithm of odds (lod) ratio scores above the chromosome (grey bar). The markers GP21 and SPUD237 delimit the QTL to an 812-kb interval. Fine mapping reduces the interval to 110 kb, a region that contains the *StCDF1* gene. **b**, Wild-type (WT) *S. andigenum*, 4 weeks after planting. **c**, *S. andigenum* transformed with *35S::StCDF1.2* at 4 weeks after planting. **d**, *S. andigenum* 14 weeks after planting. From left to right, non-transformed control and *35S::StCDF1.2* transgenic lines 7, 19 and 1. **e**, CE3027 (late maturing clone, homozygous for *StCDF1.1*) and two transgenic derivatives (CE3027 with *35S::StCDF1.2*: lines 4 and 19) 4 weeks after planting under long-day conditions. The transgenic plants show tuberization, whereas the wild type has not formed tubers. **f**, From left to right, CE3027, CE3130 (early maturing clone, containing both the *StCDF1.2* and *StCDF1.3* alleles), and CE3027 transformed with *35S::StCDF1.2* lines 4 and 19, all after 14 weeks under long-day conditions. **g**, Transgenic *S. andigenum* clone 13, showing aerial tubers. **h**, Wild-type *A. thaliana* grown under long-day conditions shown at 4 weeks after planting with flower visible (arrowhead). **i**, *A. thaliana* transformed with the *35S::StCDF1.2* construct shown at 4 weeks after planting. **j**, The same transgenic *A. thaliana* line as in (**i**), 8 weeks after planting showing a flower spike (arrowhead). Scale bars represent 1 cm (**b**, **c**, **h**–**i**) or 20 cm (**d**–**f**).

S. andigenum plants had already formed tubers 4 weeks after planting (Fig. 1c), whereas the untransformed controls are unable to tuberize under long days (Fig. 1b). Some transgenic *S. andigenum* plants also showed aerial tubers (Fig. 1g), indicative of a strong promotion of tuberization. Overexpression of the *StCDF1.2* allele also resulted in early tuber formation in the late maturing CE3027 background (Fig. 1e); these lines now tuberize at the same time as the earliest C × E genotypes (CE3130). Early maturing C × E genotypes are shorter and bushier compared to the late genotypes and this change is also observed in the *S. andigenum* and C × E transgenic clones (Fig. 1b compared to Fig. 1c, e). Furthermore, these plants also showed accelerated signs of senescence and terminated their life cycles 5 weeks earlier than the non-transgenic controls (Fig. 1d, f, respectively). Post-tuberization senescence is a normal physiological response in potato. Overexpression of the *StCDF1.2* allele thus phenocopies the early tuberization, shoot architecture and maturation traits of early genotypes. No effect of overexpression of the *StCDF1.2* gene was observed on flowering in potato, indicating a separation of signal transduction pathways for tuberization and flowering (data not shown). However, introduction of the *35S::StCDF1.2* construct into *A. thaliana* causes a late flowering phenotype (Fig. 1h–j and Supplementary Fig. 4), suggesting a conserved repressive function of potato StCDF1 on *AtCO* expression, as seen for the *A. thaliana CDF1* factor[13].

In addition to the DOF domain, CDF proteins share three conserved basic domains in their C terminus (Supplementary Fig. 2), which were reported to be important for protein–protein interaction[13]. In *A. thaliana,* blue light perception by the FKF1 LOV domain promotes binding of this protein to GIGANTEA and interaction of the FKF1-GI complex with CDF1. Ubiquitination of this DOF factor marks this transcriptional repressor for proteasomal degradation, hence facilitating *CO* expression late in the afternoon[22]. Loss of the C-terminal end in the *StCDF1.2* and *StCDF1.3* alleles may impair interaction of these truncated proteins with the FKF1–GI complex, thus leading to accumulation of these repressors and the observed effects on tuberization onset. To test this hypothesis, the *StCDF1.1*, *StCDF1.2* and *StCDF1.3* allelic variants were used as bait in a yeast two-hybrid (Y2H) assay, using the potato *StGI1* and *StFKF1* homologues as prey. Results from these interaction studies showed that while the protein encoded by the *StCDF1.1* allele strongly interacts with both StGI1 and StFKF1, the truncated proteins encoded by the early alleles (*StCDF1.2* and *StCDF1.3*) are unable to bind StFKF1, although they retain StGI1 binding affinity (Fig. 2a). These results are consistent with the putative location of a conserved binding site for StFKF1 at the C terminus of StCDF1.1 (domain III, Supplementary Fig. 2b), which is missing in the *StCDF1.2* and *StCDF1.3* alleles (Supplementary Fig. 2a, b). Furthermore, the results indicate that domain II in the StCDF1.1 protein (Supplementary Fig. 2b) is likely to be involved in StGI1 interaction. *In vivo* confirmation of these interaction results was obtained by bimolecular fluorescent complementation (BiFC) experiments, in which *Nicotiana benthamiana* leaves were co-infiltrated with split GFP constructs for the StFKF1 and the StCDF1.1 and StCDF1.2 proteins. Nuclear fluorescence of reconstituted GFP was observed only with the StFKF1 and the StCDF1.1 proteins, whereas no such interaction was found for the StCDF1.2 protein lacking C-terminal domain III (Fig. 2b, c, respectively). Additional BiFC experiments and the controls for these experiments are described in Supplementary Information and shown in Supplementary Fig. 4.

To assess whether lack of domain III leads to stabilization of the StCDF1.2 and StCDF1.3 proteins, we isolated nuclear protein extracts from leaves of plants grown in short-day conditions (06:00–14:00 light period), over a 24-hour time series. StCDF1 protein was detected by western blotting. In the late diploid potato CE3027, homozygous for the *StCDF1.1* allele, we detected a peak in protein levels between 10:00 in the morning and 14:00 in the afternoon, with much lower levels of this protein observed during the rest of the day (Fig. 2d). However, in the early maturing diploid clone CE3130 carrying the truncated SCDF1.2 and SCDF1.3 variants, a band corresponding to both proteins was present at more or less constant levels at all time-points tested (Fig. 2d). This result indicates that whereas StCDF1.1 is destabilized by StFKF1, StCDF1.2 and StCDF1.3 are not subject to this post-translational regulation and remain constant throughout the day.

©2013 Macmillan Publishers Limited. All rights reserved

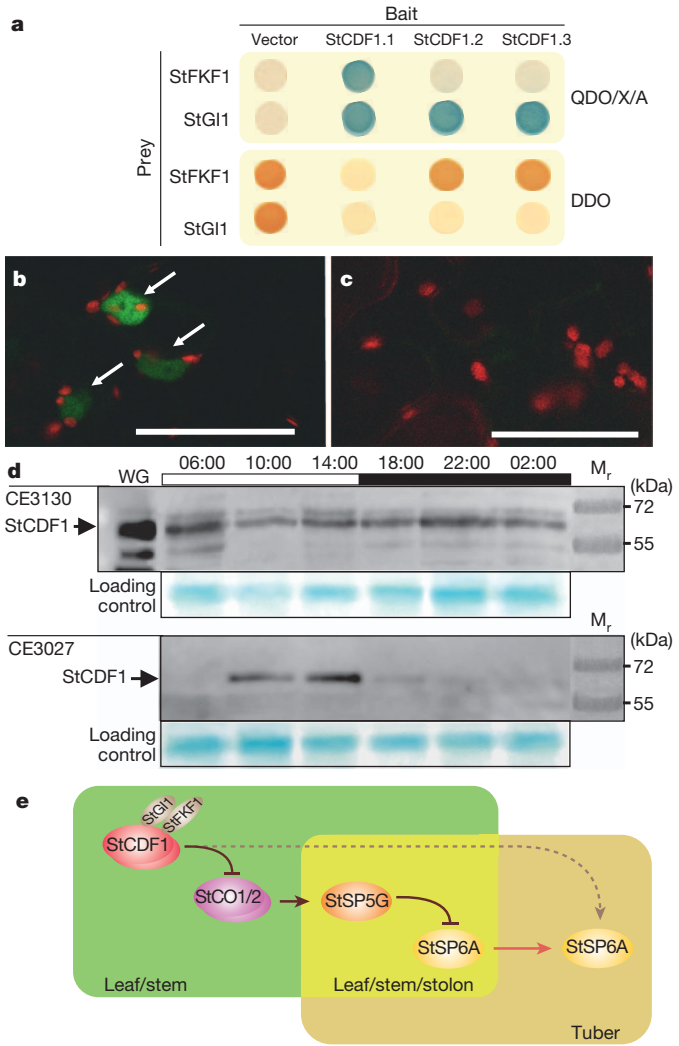


Figure 2 | Alleles of *StCDF1* exhibit different post-translational regulation throughout the day. **a**, Yeast two-hybrid assays: StCDF1.1, StCDF1.2 and StCDF1.3 proteins were bait and StFKF1 and StGI1 used as prey. Yeast was grown on QDO/X/A (top) or double dropout DDO media (bottom). **b**, BiFC analysis of StCDF1.1 and StFKF1 interaction. Arrows indicate fluorescent nuclei (scale bar, 50 μm). **c**, BiFC analysis of StCDF1.2 and StFKF1, showing no interaction. **d**, Western blot of StCDF1 protein in the clones CE3027 (late maturing, homozygous for the *StCDF1.1* allele) and CE3130 (early maturing, containing both the *StCDF1.2* and *StCDF1.3* alleles) under short days. Leaf nuclear protein was harvested at 6 time points. The bar indicates light and dark period. Lane 1 (top panel) is StCDF1.1 protein obtained in a wheat-germ system. Equal levels of nuclear leaf protein were loaded in each lane. Coomassie blue stained loading controls below. The right lane shows molecular weight markers (72 and 55 kilodalton (kDa) bands). **e**, Model for the regulation of tuber development. Green represents aerial plant organs, the tuber is shown in brown and the overlap represents leaf, stem and stolon. The lines represent induction (→) or repression (⊣). The red arrow represents transport. StCDF1 acts as an indirect inducer of *StSP6A* (dotted arrow).

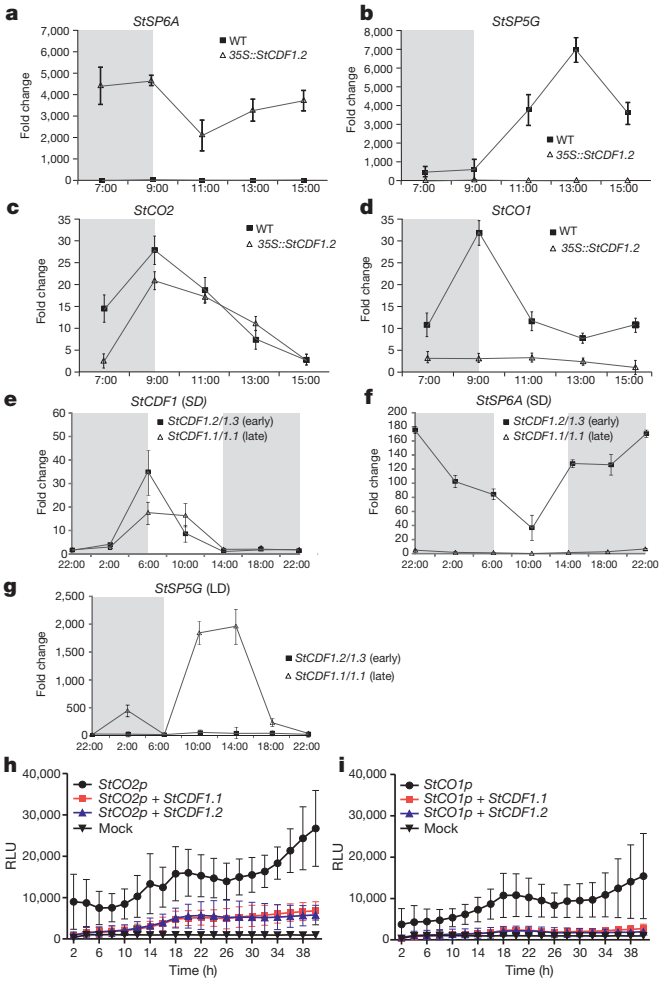


Figure 3 | Expression analysis of genes regulating tuber formation in *S. andigenum*, late and early C × E descendants, and transient co-expression analysis in *N. benthamiana* leaves. **a**–**d** The mean of three biological replicates were taken for gene expression analysis of *S. andigenum* wild type (square) and transgenic plants carrying the *StCDF1.2* overexpression construct (*35S::StCDF1.2*, triangle). Five time points for leaf harvest (time is shown on the *x* axis) from plants grown in long days. Expression of *StSP6A* (**a**), *StSP5G* (**b**), *StCO2* (**c**) and *StCO1* (**d**). **e**–**g**, *StCDF1* (**e**) and *StSP6A* (**f**) mRNA levels were measured in short-day (SD) conditions in two diploid potato genotypes. The mRNA of *StSP5G* (**g**) was measured in long-day (LD) conditions. Seven time points over 24 h were analysed from genotypes that are either early maturing heterozygotes carrying the *StCDF1.2* and *StCDF1.3* alleles (*StCDF1.2/1.3*, square) or late maturing homozygotes carrying two copies of the *StCDF1.1* allele (*StCDF1.1/1.1*, triangle). Error bars represent the standard error of 3 biological replicates. In **a**–**g**, the *y* axis represents fold change compared to the lowest observed value in the series and the shaded area of the graph represents the dark period. **h**, **i**, Relative luminscence units (RLU) from transient expression studies of the *35S::StCDF1.1* or *35S::StCDF1.2* effector constructs co-transformed with the *StCO2* and *StCO1* promoters fused to the *LUC* reporter gene (*StCO2p* and *StCO1p*). *Agrobacterium* with these constructs were co-infiltrated into *N. benthamiana* leaves and luciferase activity was measured at intervals of 2 h for 2 days. Error bars represent the standard deviation of 4 biological replicates.

High level of *StCDF1.2* expression conferred by the CaMV *35S* promoter was verified in the transgenic *S. andigenum* lines (Supplementary Fig. 5a). As these transgenic plants tuberize at an early stage under long-day conditions, we also followed the expression of the potato *FT* homologue *StSP6A* throughout the day. As expected, this gene was strongly upregulated compared to non-transgenic wild-type controls (Fig. 3a). In contrast, a second *FT* homologue, *StSP5G* is downregulated in the transgenic lines but peaks in the controls just after noon (Fig. 3b). CDF proteins were reported in *A. thaliana,* to directly bind to the *CO* promoter and interfere with expression of this gene[13]. In potato, two CONSTANS genes (*StCO1* and *StCO2*; hereafter *StCO1/2* refers to both genes together) were identified, and activity of these genes was shown to repress tuber formation in long days[16] (Supplementary Information). It is noteworthy that both genes and in particular *StCO1* (Fig. 3d) are transcriptionally repressed in the transgenic *S. andigenum* lines harbouring the *35S::StCDF1.2* transgene under long-day conditions (Fig. 3c, d), indicating a conserved repressive function of potato StCDF1 on StCO1/2. Although the suppressive function of StCO1/2 on tuberization was not previously well understood, more recent evidence suggests that these factors, rather than

©2013 Macmillan Publishers Limited. All rights reserved

directly regulating *StSP6A* expression are instead implicated in *StSP5G* activation in long days. This *FT* homologue appears to function as a suppressor of *StSP6A* gene expression, similar to the way that the sugar beet *BvFT1* homologue represses flowering by antagonising expression of the floral promoter *BvFT2* gene[9]. Hence, these observations are consistent with a model in which repression of *StCO1/2* expression by *StCDF1* indirectly induces *StSP6A* expression by downregulating the *StSP5G* gene. These results are also consistent with an independent study in which overexpression of the *A. thaliana LKP2* gene in potato was found to result in an early tuberization phenotype[23].

To gain insight into the diurnal regulation of *StCDF1*, *StCO1, StCO2* and *StSP6A* genes in a non-transgenic background, we analysed expression of these genes in the early and late C × E genotypes. We found that *StCDF1* expression peaks just before dawn under short-day conditions, and that transcript levels for this gene are higher in the early genotypes carrying the *StCDF1.2* and *StCDF1.3* allele combination (Fig. 3e). Early genotypes also display strong upregulated expression of the *StSP6A* gene under inductive short-day conditions, compared to genotypes that are homozygous for *StCDF1.1* and confer a late phenotype (Fig. 3f). Under short-day conditions, no *StSP5G* transcript could be detected in late or early genotypes (data not shown). However, under long days *StSP5G* transcript is highly expressed, particularly in late genotypes homozygous for *StCDF1.1* compared to the relatively low levels detected in the *StCDF1.2* and *StCDF1.3* descendants, with an early maturing genotype (Fig. 3g). These results indicate that StCDF1 acts as an indirect inducer of *StSP6A* by downregulating the *StSP6A* repressor *StSP5G* through *StCO1/2* (Fig. 2c). These gene expression profiles in early C × E genotypes mirror the situation we find in transgenic plants that overexpress *StCDF1.2*.

In an extended analysis of both upstream and downstream genes associated with *StSP6A* regulation, we did not find any transcriptional effects on genes directly involved in the regulation of the circadian clock (data not shown). However, we did find evidence for negative feedback regulation of the genes directly binding to and regulating StCDF1 protein levels, such as *StGI1* and *StFKF1*, in addition to forward regulation of genes involved in alternative routes of tuber development such as *StBEL5* (Supplementary Fig. 5b–d). This indicates a possible broader regulatory function of the StCDF1 protein in potato than just a role in *StCO1/2* repression.

To find out if there is a transcriptional regulatory link between *StCDF1* and *StCO1/2* genes, the *35S::StCDF1* construct was co-infiltrated into tobacco leaves with *StCO1* and *StCO2* promoter constructs driving expression of the firefly luciferase gene. The results show that both *StCDF1.2* and *StCDF1.3* suppress expression of either of the *StCO* genes (Fig. 3h, i), with a stronger suppression of the *StCO1* promoter construct. These results are in line with the gene expression data described here and show that truncation of the C-terminal end of the StCDF1 proteins stabilizes these factors but does not impair their DNA binding ability, hence providing strong evidence for a linkage association between *StCDF1* and *StCO1/2* expression. Electrophoretic mobility shift assay (EMSA) studies with the StCDF1.1 and StCDF1.2 proteins further confirm that these proteins bind with similar affinities to elements in the *StCO1/2* promoters, thus corroborating the results obtained from *in vivo* repression assays (Supplementary Information and Supplementary Fig. 6).

The major role of the *StCDF1* gene for plant maturity was verified in the wider gene pool of cultivated potato. A genome-wide association study, using a representative set of 83 potato cultivars, allowed the identification of highly significant marker-by-trait associations for single nucleotide polymorphism (SNP) markers within and in short-range linkage disequilibrium with the causal insertion polymorphisms in *StCDF1* (Supplementary Information and Supplementary Fig. 7). Beyond this chromosome 5 region, no other genomic regions showed strong association with plant maturity. This suggests that putative allelic forms of all other genes involved in the signalling cascade do not cause phenotypic variation for plant maturity.

Based on our results and the currently available knowledge of tuber development in potato, we propose a model for the induction of tuberization in which the clock output proteins StGI and StFKF1 regulate the abundance of the transcriptional regulator StCDF1. StCDF1 downregulates *StCO1/2,* which supresses the transcription of *StSP5G*, enabling expression of the mobile StSP6A signal and resulting in the induction of tuber development at the stolon termini. The allelic variation uncovered in this work is likely to have formed the basis of the domestication of potato in latitudes characterized by large summer/winter day-length variation and may in future allow tailored breeding for new geographic regions.

METHODS SUMMARY

Plant materials and growth conditions. *Solanum tuberosum* clones (CE3130, CE3027, *S. andigenum* 7540 and transgenic derivatives) and *Arabidopsis thaliana* ecotype Col-0 were greenhouse grown under long days. For tuberization, clones were transferred to chambers with an 8 h light and 16 h dark regime. Transgenic plants were generated using *Agrobacterium tumefaciens* transformation of inter-node explants. *A. thaliana* transgenic lines were generated using the floral dipping method.
Plasmid constructs. *StCDF1* coding region was cloned from complementary DNA into the pK7WG2 plasmid to generate overexpression lines. Other constructs for the yeast two-hybrid analysis (*StGI* and *StFKF1*) were generated according to the manufacturer's recommendations using the Matchmaker Gold kit (Clontech).
Real-time PCR analyses. Expression analysis was carried out using 1 μg of total RNA isolated from leaves of plants grown under controlled conditions. Real-time gene expression was analysed on a Biorad C1000 thermocycler using iQ-SYBR-green supermix (Bio-Rad). For potato the *EIF3e* gene was used as a control (RT_REF primers; Supplementary Table 1).

Full Methods and any associated references are available in the online version of the paper.

Received 17 July 2012; accepted 15 January 2013.
Published online 6 March 2013.

1. Yanagisawa, S. Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant Cell Physiol.* **45,** 386–391 (2004).
2. Navarro, C. *et al.* Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature* **478,** 119–122 (2011).
3. Spooner, D. M., McLean, K., Ramsay, G., Waugh, R. & Bryan, G. J. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proc Natl Acad. Sci. USA* **102,** 14694–14699 (2005).
4. Salaman, R. N. *The History and Social Influence of the Potato* (Cambridge Univ. Press, 1989).
5. Bäurle, I. & Dean, C. The timing of developmental transitions in plants. *Cell* **125,** 655–664 (2006).
6. Koornneef, M., Hanhart, C. J. & van der Veen, J. H. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **229,** 57–66 (1991).
7. Kojima, S. *et al. Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant Cell Physiol.* **43,** 1096–1105 (2002).
8. Lifschitz, E. & Eshed, Y. Universal florigenic signals triggered by *FT* homologues regulate growth and flowering cycles in perennial day-neutral tomato. *J. Exp. Bot.* **57,** 3405–3414 (2006).
9. Pin, P. A. *et al.* An antagonistic pair of *FT* homologs mediates the control of flowering time in sugar beet. *Science* **330,** 1397–1400 (2010).
10. Suárez-López, P. *et al. CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis. Nature* **410,** 1116–1120 (2001).
11. Yanovsky, M. J. & Kay, S. A. Molecular basis of seasonal time measurement in *Arabidopsis. Nature* **419,** 308–312 (2002).
12. Park, D. H. *et al.* Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis GIGANTEA* gene. *Science* **285,** 1579–1582 (1999).
13. Imaizumi, T., Schultz, T. F., Harmon, F. G., Ho, L. A. & Kay, S. A. FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in *Arabidopsis. Science* **309,** 293–297 (2005).
14. Sawa, M., Nusinow, D. A., Kay, S. A. & Imaizumi, T. FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis. Science* **318,** 261–265 (2007).
15. Song, Y. H., Smith, R. W., To, B. J., Millar, A. J. & Imaizumi, T. FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. *Science* **336,** 1045–1049 (2012).
16. González-Schain, N. D., Diaz-Mendoza, M., Zurczak, M. & Suarez-Lopez, P. Potato CONSTANS is involved in photoperiodic tuberization in a graft-transmissible manner. *Plant J.* **70,** 678–690 (2012).
17. Rodríguez-Falcón, M., Bou, J. & Prat, S. Seasonal control of tuberization in potato: conserved elements with the flowering response. *Annu. Rev. Plant Biol.* **57,** 151–180 (2006).

©2013 Macmillan Publishers Limited. All rights reserved

18. Visker, M. H. *et al.* Can the QTL for late blight resistance on potato chromosome 5 be attributed to foliage maturity type? *Theor. Appl. Genet.* **106,** 317–325 (2003).
19. The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature* **475,** 189–195 (2011).
20. Abelenda, J. A., Navarro, C. & Prat, S. From the model to the crop: genes controlling tuber formation in potato. *Curr. Opin. Biotechnol.* **22,** 287–292 (2011).
21. Martínez-García, J. F., Virgos-Soler, A. & Prat, S. Control of photoperiod-regulated tuberization in potato by the *Arabidopsis* flowering-time gene *CONSTANS*. *Proc. Natl Acad. Sci. USA* **99,** 15211–15216 (2002).
22. Takase, T. *et al. LOV KELCH PROTEIN2* and *ZEITLUPE* repress *Arabidopsis* photoperiodic flowering under non-inductive conditions, depending on *FLAVIN-BINDING KELCH REPEAT F-BOX1*. *Plant J.* **67,** 608–621 (2011).
23. Inui, H., Ogura, Y. & Kiyosue, T. Overexpression of *Arabidopsis thaliana LOV KELCH REPEAT PROTEIN 2* promotes tuberization in potato (*Solanum tuberosum cv.* May Queen). *FEBS Lett.* **584,** 2393–2396 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank K. Kaufman, G. Angenent, N. Mansoori Zangir, J. Uitdewilligen and R. Hutten for help and advice in the planning and execution of the project. Thanks also go to V. Ghasemi Omran and I. Pereira-Bertram for their expert technical assistance and to T. Bisseling for critical reading of the manuscript. We gratefully acknowledge the Dutch potato breeding companies, Argrico Research B.V., HZPC Holland B.V. and Meijer B.V. for the execution of the field trials. The project was supported by the Technology Foundation STW (grant no. WGC.7795 and WPB.5283) and FP6 integrated Project EU-Sol PL016214. M.d.M.C.G. received additional support from Almeria University Fund.

Author Contributions B.K., J.A.A., M.d.M.C.G., M.O., J.M.d.B., K.K., B.M.H., C.S. and C.W.B.B. performed the experiments and analysed the results. B.K., S.P., H.J.v.E., J.M.d.B., R.G.F.V. and C.W.B.B. designed the experiments, C.W.B.B. wrote the manuscript and J.A.A., S.P., B.M.H., B.K., M.d.M.C.G., H.J.v.E., J.M.d.B., K.K. and R.G.F.V. gave critical feedback on the manuscript.

Author Information The sequence of the *StCDF1.2* gene variant is available in GenBank (accession number KC284850). All other sequences from potato are available at GenBank or the Potato Genome Sequencing Consortium website (http://www.potatogenome.net). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.W.B.B. (Christian.Bachem@wur.nl).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Plant material and growth conditions. The plant material used for genetic and phenotypic analysis was obtained from the parents and progeny of two diploid potato-mapping populations. The cross between the parental genotypes SH82-93488 (SH) and RH89-039-16 (RH) yielded the F_1 mapping population of the ultra-dense amplified fragment length polymorphism (AFLP) marker genetic map of potato (Supplementary Information). The backcross between genotype C and its F_1 descendant E, resulted in the C × E mapping population for which an AFLP marker linkage map is available that has recently been expanded with SNP markers (Supplementary Information). Both core mapping populations and their parents are maintained by annual propagation of tubers. In addition, new seedlings were raised from these crosses to use for fine mapping of the plant maturity phenotype. Both these populations segregate for plant maturity (Supplementary Information). This trait has many pleiotropic affects such as development of canopy, vegetative growth, onset of tuberization, leaf senescence, life cycle length and pathogen resistance[18].

For transformed and non-transformed controls, the genotype *S. tuberosum* group *andigenum* line 7540 (ref. 2) was grown in a greenhouse in long-day conditions (cycling 16 h light period followed by 8 h dark period, referred to as long day). This genotype does not form tubers under these light cycling conditions and is dependent on a long dark period for the initiation of tuber formation. *Agrobacterium* mediated transformation using stem internodes was carried out as described previously (Supplementary Information). Progeny from the C × E population were phenotyped for late and early maturing genotypes were determined as described subsequently. Clone CE3130 (used as a recipient for the *35S::StCDF1.2* construct) shows an extremely early maturity phenotype, whereas CE3027 is a very late maturing descendant. These early and late C × E descendants were used in real-time PCR experiments to determine gene expression levels as described here. Experiments to determine the tuberization phenotype were carried out three times and gave consistent results (Supplementary Information). Isolation of RNA and protein from the C × E genotypes CE3130 (early maturing) and CE3027 (late maturing) was carried out using protocols described here. Sprouting tubers from these genotypes were planted in 20 cm pots with compost in light-, temperature- and humidity-controlled growth chambers for 3 weeks under long-day conditions (16 h light (20°C); 8 h dark (18°C); 75% average relative humidity). The plants were then switched to short-day conditions (8 h light, 16 h dark) for two weeks before leaf harvest. Leaf material from the fourth and fifth node (from the apex) was harvested in two biological replicates. Leaf material from 5 plants was pooled for each replicate. Materials were harvested at 6 time points over a 24-h period (at 2:00, 6:00, 10:00, 14:00, 18:00 and 22:00). For gene expression analysis of the transformed *S. andigenum* plants and the non-transformed controls, plants were also grown in growth chambers (already described here for the C × E genotypes). Leaf RNA was isolated from plants grown under both long-day and short-day conditions as specified in the results sections. Leaf material was either harvested at the same 6 time points as described here or when a single measurement is indicated in the results, leaves from the 10:00 and 18:00 harvests were used.

For transformation of the *A. thaliana* ecotype Columbia (Col-0), a floral dip protocol was used and untransformed Col-0 plants were used as controls (Supplementary Information). Plants were grown in the greenhouse under long-day conditions. Plants were checked for transition to flowering at 3-day intervals. RNA was isolated from leaf material for gene expression analysis at time points indicated in the results using the same protocols as described for potato here.

Cloning of the St*CDF1* gene and construction of the *35S::StCDF1.2* plasmid. The *StCDF1* gene was cloned from complementary DNA (cDNA) of the early CE3130 clone using specific primers from the *StCDF1.2* allele (forward: 5′-ATG TCTGAAGTTAGAGATCCTGCTAT-3′, reverse: 5′-CTCTCGCGGAAATGAA GTGA-3′). Cloning was carried out according to the manufacturer's recommendations using the Gateway cloning system into the pENTR/D-TOPO cloning kit (Invitrogen). The LR clonase was used to insert the gene into the pK7WG2 binary vector[24] and this construct was transferred into *Agrobacterium tumefaciens* for transformation of potato and *A. thaliana*.

DNA isolation and sequencing from tetraploid genotypes. Leaf material from the tetraploid potato cultivars was harvested and DNA isolated using the high-throughput DNA extraction method as described previously[25]. Genome-wide association studies were carried out as described in the Supplementary Information.

Plant maturity phenotyping. Eighty F_1 genotypes from the ultra-dense SH × RH genetic map were evaluated for their maturity type in two field trials on clay soil during summer 2004. Field trials were performed by Agrico Research B.V. in Bant (52° 46′ N, 5° 45′ E), The Netherlands and by Meijer B.V. in Rilland (51° 25′ N, 4°11′ E), The Netherlands. For each genotype, eight tubers were planted per trial field, in a duplicate block design. The plant maturity score of the F_1 plants was determined three months after planting from the overall status of above-ground plant development, by comparison to the maturity status of eight reference potato varieties, and is expressed as a score between 1 (very late) and 9 (very early). Relatively late maturing reference varieties were Agria (5) and Desiree (5.5). Intermediate varieties were Santana (6), Lady Rosetta (6.5) and Bintje (6.5). Relatively early maturing was Monalisa (7.5) and early reference varieties were Premiere (8.5) and Agata (8.5). The maturity scores were used for QTL analysis on the separate maternal (SH) and paternal (RH) AFLP genetic maps by interval mapping (Supplementary Information).

Newly produced siblings from the SH × RH and C × E mapping populations displaying recombination events in the GP186–GP179 genetic marker interval were multiplied *in vitro* by stem cuttings in order to generate sufficient material for phenotyping 4 hill plots in two replicates in Wageningen (51° 58′ N, 5° 40′ E) in May 2008. Plant maturity type values were scored in August 2008. A subset of the initial sibling collection, with recombination events close to the maturity locus, was propagated in large pots in open-air screen cages during 2009. The tubers from this multiplication were used to repeat the maturity phenotyping of these recombinant plants in a field experiment in Wageningen in the summer of 2010 (Supplementary Fig. 1).

Marker screening and recombinant analysis. The initial recombinant analysis of the plant maturity trait was done with cleaved amplified polymorphic sequences (CAPS) markers for standard genetic loci on the north arm of potato chromosome 5 (GP186, GP21, SPUD238 and GP179; Supplementary Information). Locus GP179 was not polymorphic in the C × E population and a substitute marker was developed from the sequence of chromosome-5 BAC clone RH185H09, 40 kb south of marker GP179. Further fine mapping of the plant maturity locus in the GP21–SPUD237 interval involved PCR markers, which were developed from sequenced BAC clones from the RH parent in this region. The PCR markers were designed to give polymorphisms that can be mapped on the parental genetic maps of RH and E, so that they would provide the haplotype information needed for fine mapping of the plant maturity trait. The PCR markers were examined by CAPS digestion or genotyped by high resolution DNA melting (Supplementary Information).

Analysis of gene expression. To quantify relative expression levels of genes involved in photoperiodic response in potato, total RNA was isolated from 10-day-old whole leaves using the RNeasy plant mini kit (Qiagen) and DNase I (Amplification Grade, Invitrogen) digestion was performed to eliminate DNA contamination. For synthesis of cDNA, 1 μg of total RNA was reverse transcribed with the iScript cDNA Synthesis Kit (Bio-Rad) with oligo(dT)15 as a primer. cDNA was diluted with water to a volume of 400 μl, and 4 μl of diluted cDNA was used for subsequent real-time PCR reactions. Amplified products were detected using iQ-SYBR-green supermix (Bio-Rad) using a C1000 (Bio-Rad) thermal cycler. The following program was used for PCR: 95°C for 3 min, 40 cycles of 95°C for 15 s, 60°C for 1 min, followed by 95°C for 10 s, 65°C for 5 s and 95°C for 5 s. Identical procedures were used for estimating relative expression levels of genes involved in photoperiodic response in *A. thaliana*. Based on expression profiling of multiple circadian time courses, we selected the *StElF3e* gene and the isopentenyl pyrophosphate: dimethyllallyl pyrophosphate isomerase 2 (*IPP2*) gene as internal controls in potato and *A. thaliana*, respectively. The primer pairs used for expression analyses are given in Supplementary Table 1. Each PCR was repeated at least twice. All gene expressions are means ± s.e.m. of relative values of signals, normalized to the *St349* or *AtIPP2* controls from two independent experiments.

Transient assays in tobacco. Promoter regions for the *S. andigenum StCO2* (PGSC003DMT400026065) and *StCO1* (PGSC003DMT400026068) genes were amplified with primers CO_2_PROM_For (5′-CACCTTAAAAATTCAGAGA AGA-3′) CO_2_PROM_Rev (5′-ACTGACTAATTTTGGTTTCTCT-3′); and CO_1_PROM_For (5′-CACCATGGAGCAATTCTTGTGAT-3′) CO_1_PROM_ Rev (5′-TTCTTCTTGCTCTCTATTTTCTG-3′), respectively. The PCR products were cloned into the pENTR/D-TOPO vector (Invitrogen) and inserted by LR-reaction (Invitrogen) into the binary vector Luc+ Trap3 (ref. 26). These constructs were transformed into the *Agrobacterium tumefaciens* (GV3101) strain and infiltrated into *Nicotiana benthamiana* leaves together with *35S::CDF1.1* and *35S::CDF1.2* constructs. The p19 protein was used to suppress gene silencing[27]. *Agrobacterium* strains containing the *StCO1* or *StCO2* promoter::*LUC* fusions, the *35S::CDF1.1* or *35S::CDF1.2* constructs and the p19 silencing plasmid at an attenuance $D_{600\,nm}$ of 0.7, 0.7 and 1.0, respectively, were resuspended in buffer (10 mM MES, pH 5.5, 10 mM $MgSO_4$ and 150 μM acetosyringone), and used for infiltration. Two days after infiltration leaf disks were incubated in 2× MS supplemented with 0.02 mg ml^{-1} D-luciferin (Sigma-Aldrich) and luciferase activity recorded every 2 h with a microplate luminometer LB-960 (Berthold

©2013 Macmillan Publishers Limited. All rights reserved

Technologies) and further analysed with the Mikrowin 2000 (version 4.29) software (Mikrotek Laborsysteme).

Yeast two-hybrid analysis. We used the GAL4-based yeast two-hybrid system (Matchmaker Gold yeast two-hybrid system, Clontech) to analyse the protein–protein interactions in yeast. The cDNAs encoding the full-length *StCDF1.1*, *StCDF1.2* and *StCDF1.3* alleles were amplified using specific forward primers containing the EcoRI site and reverse primers containing the BamHI site (see Supplementary Table 1) and were cloned into the EcoRI and BamHI sites of pGBKT7 for bait constructs. Prey vectors comprising the coding region for *StFKF1* and *StGI1* genes were constructed by cloning the full length *StFKF1* (PGSC0003DMG400019971) and *StGI* (PGSC0003DMG200025064) cDNA amplified by specific forward primers containing the EcoRI site and reverse primers containing the BamHI site, and cloned into EcoRI and BamHI sites of the prey vector pGADT7. The bait and prey clones were verified by sequencing. We used an empty pGADT7 plasmid as a negative control (prey) for yeast two-hybrid analysis. All baits used in these experiments (*StCDF1.1, StCDF1.2* and *StCDF1.3*) tested negative for auto-activation according to the manufacturer's protocol. The haploid yeast strains harbouring each bait vector (Y2H Gold strain) and each prey vector (Y187 strain) were mated on the 2× YPDA medium at 30°C overnight. The diploid yeast cells that carry both bait and prey vectors were selected on both SD/−Leu/−Trp double dropout medium (DDO; for selection of bait and prey vectors), and SD/−Ade/−His/−Leu/−Trp/−X-α-Gal/AbA, Quadruple Dropout medium (QDO/X/A; for selection of both vectors and the interaction selection) plates to score the protein–protein interaction by growth. After a 3-day incubation at 30°C, images of the colonies on both plates were recorded.

Preparation of plant protein extracts and immunoblot analysis. Nuclear extracts were prepared essentially as described[27]. Total nuclear protein content was determined photometrically by Bradford assay (Bio-Rad protein assay), samples were normalized for equal protein content and 10 μg of nuclear protein was loaded per sample. *In vitro* transcription/translation of *StCDF1.1* allele coding sequence was performed using the TNT coupled wheat germ extract transcription/translation system (Promega) according to the manufacturer's instructions, in order to identify the StCDF1 protein by size. Nuclear extracts and wheat germ synthesized proteins were separated on 12% SDS–PAGE gels and transferred to nitrocellulose membranes. After transfer, SDS–PAGE gels were stained with colloidal Coomassie blue. A protein band of unknown identity that had constant staining in all samples tested was chosen as a control for equal loading verification. StCDF1 proteins were detected using affinity-purified rabbit anti-StCDF1 antibody. The antibody was raised against a synthetic peptide (YLHDDHTTSSPLSIDD, highlighted in green in Supplementary Fig. 2) unique to StCDF1 and affinity purified according to the manufacturer's protocol (Eurogentec). At the primary antibody reaction step, the affinity-purified StCDF1 antibody was diluted 1:250 in Tris-buffered saline containing 0.05% Tween 20 in 1× TBST and 3% non-fat dry milk, followed by 2 h incubation with Goat peroxidase–conjugated anti-rabbit IgG (Rockland) secondary antibody diluted 1:15,000 times in Tris-buffered saline containing 0.05% Tween 20 in 1× TBST and 3% non-fat dry milk. Immunoreactive proteins were visualized by a reaction mixture with stable peroxide solution and luminol/enhancer solution. Signals were recorded by exposure to G-Box camera system (Syngene).

Biomolecular fluorescence complementation (BiFC) analysis. Full-length open reading frames for the *Solanum tuberosum StFKF1*, *StGI* and *StCDF1.1* and *StCDF1.2* genes were cloned in pENTR/D-TOPO vector (Invitrogen) and inserted by LR-reaction (Invitrogen) into the binary vectors pYFN43 and pYFC43 (provided by A. Ferrando), containing the N- or C-terminus of the GFP protein, respectively. Constructs were transformed into *Agrobacterium tumefaciens* (GV3101) and co-infiltrated into the abaxial surface of 2–3-week-old *Nicotiana benthamiana* plants together with the p19 plasmid of tomato bushy stunt virus to suppress gene silencing[27]. After two days of infiltration, leaves were treated with the proteasome inhibitor MG132 at 50 μM. Plants were then kept under blue light (fluence rate of 19 $\mu mol\, m^{-2}\, s^{-1}$) to promote StFKF1–StGI interaction, before confocal microscopy observation. Plants were monitored using a Leica TCS SP5 laser-scanning confocal microscope.

Electrophoretic mobility shift assay (EMSA). Full-length open reading frames for the *Solanum tuberosum StCDF1.1* and *StCDF1.2* alleles were cloned into the pMal-C2X vector, to generate fusions of these proteins with the maltose-binding protein (MBP). MBP–StCDF1.1 and MBP–StCDF1.2 proteins were purified from bacteria with amylose resin beads. Double-stranded biotinylated probes including theoretical StCDF1 binding sites on the *StCO2* (PGSC003DMT400026065) and *StCO1* (PGSC003DMT400026068) promoters were generated by amplification of the oligonucleotide templates: *StCO2:* 5′-GATCCTCTCGCGTACAATAAAAAAGTTGTAAAGTTTGAGACGTAGCGTATGCGTA-3′ and *StCO1* 5′-GATCCTCTCGCGTACAAAAAAGTGAAGTTATTCGGAAAAGCGTAGCGTATGCGTA-3′, with the biotinylated primers 5′-GATCCTCTCGCGTAC-3′ and 5′-TACGCATACGCTACG-3′. For the EMSA reaction both proteins were incubated with the biotinylated probes in binding buffer (10 mM Tris, pH 7.5, 50 mM KCl, 1 mM dithiothreitol, 5 mM $MgCl_2$, 50 ng μl^{-1} poly(dI-dC)) for 15 min at room temperature separated on 6% PAGE gels in 0.5× TBE buffer. A 10- and 100-fold excess of unlabelled probes were used for specific competition. After semidry transfer to a nitrocellulose membrane, the LightShift chemiluminescent EMSA kit (Pierce) was used in all detection steps.

24. Karimi, M., Inzé, D. & Depicker, A. GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7,** 193–195 (2002).
25. Gady, A. L. F. *et al.* Implementation of two high through-put techniques in a novel application: detecting point mutations in large EMS mutated plant populations. *Plant Methods* **5,** 13 (2009).
26. Calderon-Villalobos, L. I. *et al.* LucTrap vectors are tools to generate luciferase fusions for the quantification of transcript and protein abundance *in vivo*. *Plant Physiol.* **141,** 3–14 (2006).
27. Voinnet, O., Rivas, S., Mestre, P. & Baulcombe, D. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33,** 949–956 (2003).

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature12005

Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC

V. Stalin Raj[1]*, Huihui Mou[2]*, Saskia L. Smits[1,3], Dick H. W. Dekkers[4], Marcel A. Müller[5], Ronald Dijkman[6], Doreen Muth[5], Jeroen A. A. Demmers[4], Ali Zaki[7], Ron A. M. Fouchier[1], Volker Thiel[6,8], Christian Drosten[5], Peter J. M. Rottier[2], Albert D. M. E. Osterhaus[1], Berend Jan Bosch[2] & Bart L. Haagmans[1]

Most human coronaviruses cause mild upper respiratory tract disease but may be associated with more severe pulmonary disease in immunocompromised individuals[1]. However, SARS coronavirus caused severe lower respiratory disease with nearly 10% mortality and evidence of systemic spread[2]. Recently, another coronavirus (human coronavirus-Erasmus Medical Center (hCoV-EMC)) was identified in patients with severe and sometimes lethal lower respiratory tract infection[3,4]. Viral genome analysis revealed close relatedness to coronaviruses found in bats[5]. Here we identify dipeptidyl peptidase 4 (DPP4; also known as CD26) as a functional receptor for hCoV-EMC. DPP4 specifically co-purified with the receptor-binding S1 domain of the hCoV-EMC spike protein from lysates of susceptible Huh-7 cells. Antibodies directed against DPP4 inhibited hCoV-EMC infection of primary human bronchial epithelial cells and Huh-7 cells. Expression of human and bat (*Pipistrellus pipistrellus*) DPP4 in non-susceptible COS-7 cells enabled infection by hCoV-EMC. The use of the evolutionarily conserved DPP4 protein from different species as a functional receptor provides clues about the host range potential of hCoV-EMC. In addition, it will contribute critically to our understanding of the pathogenesis and epidemiology of this emerging human coronavirus, and may facilitate the development of intervention strategies.

Coronaviruses infect a wide range of mammals and birds. Their tropism is primarily determined by the ability of the spike (S) entry protein to bind to a cell surface receptor. Coronaviruses have zoonotic potential due to the adaptability of their S protein to receptors of other species, most notably demonstrated by SARS-CoV[6], the causative agent of the SARS epidemic, which probably originated from bats[7]. Recently, a new coronavirus, named hCoV-EMC, has been identified in 13 patients to date with seven fatalities that suffered from severe respiratory illness, in some cases accompanied with renal dysfunction[3,4,8]. Genetically the virus is similar to bat coronaviruses HKU4 and HKU5 and—based on phylogenetic analysis using a small fragment of the virus—to a bat coronavirus found in *Pipistrellus pipistrellus* in the Netherlands[5]. In recent years molecular surveillance studies revealed the existence of at least 60 novel bat coronaviruses, including some closely related to SARS-CoV[9]. As the human cases do not seem to originate from transmission from one source, the epidemiology of hCoV-EMC may be explained by a concealed circulation in the human population or by the repeated introduction from an intermediate animal host. For a better understanding of the biology of this novel coronavirus, timely identification of the receptor could reveal important clues to its zoonotic transmission potential and pathogenesis and to the design of possible intervention strategies. Two types of coronavirus protein receptors have been identified: the Betacoronavirus mouse hepatitis virus uses immunoglobulin-related carcinoembryonic antigen-related cell adhesion molecules (CEACAM)[10] to enter cells, whereas for several Alpha- and Betacoronaviruses, two peptidases have been identified as receptors (aminopeptidase N (APN, CD13) for hCoV-229E and several animal coronaviruses[11,12], and angiotensin

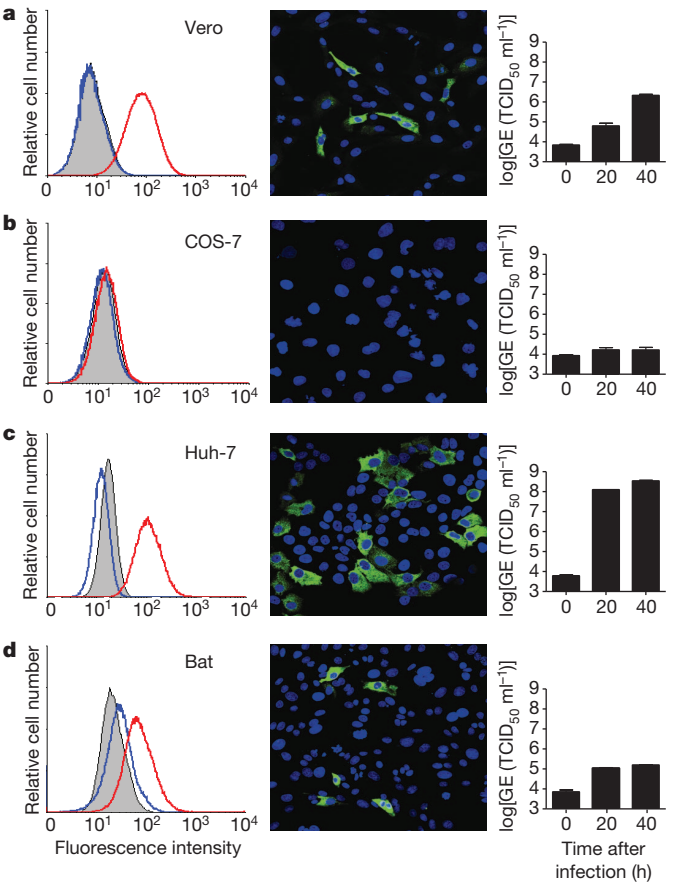


Figure 1 | Binding of hCoV-EMC S1 to cells is correlated with infection of hCoV-EMC. a–d, Shown in the left panels are the FACS analyses of hCoV-EMC S1–Fc binding (red line) to Vero (**a**), COS-7 (**b**), Huh-7 (**c**) and bat cells (**d**). A feline CoV S1–Fc protein (blue line) and mock-incubated cells (grey shading) were used as controls. In the middle panels, hCoV-EMC-infected cells are visualized using an antiserum that recognizes the non-structural protein NSP4. In the right panels, hCoV-EMC RNA levels in supernatants of the infected cells at 0, 20 and 40 h after infection were quantified using a TaqMan assay and expressed as genome equivalents (GE; half-maximal tissue-culture infectious dose ($TCID_{50}$) per ml). Error bars indicate s.e.m.

[1]Department of Viroscience, Erasmus Medical Center, 3000 CA Rotterdam, The Netherlands. [2]Virology Division, Department of Infectious Diseases & Immunology, Faculty of Veterinary Medicine, Utrecht University, 3508 TD Utrecht, the Netherlands. [3]Viroclinics Biosciences BV, 3029 AK Rotterdam, The Netherlands. [4]Proteomics Department, Erasmus Medical Center, 3000 CA Rotterdam, The Netherlands. [5]Institute of Virology, University of Bonn Medical Centre, 53105 Bonn, Germany. [6]Institute of Immunobiology, Kantonal Hospital St Gallen, 9007 St Gallen, Switzerland. [7]Virology Laboratory, Dr Soliman Fakeeh Hospital, Jeddah, Saudi Arabia. [8]Vetsuisse Faculty, University of Zürich, 8057 Zürich, Switzerland.
*These authors contributed equally to this work.

©2013 Macmillan Publishers Limited. All rights reserved

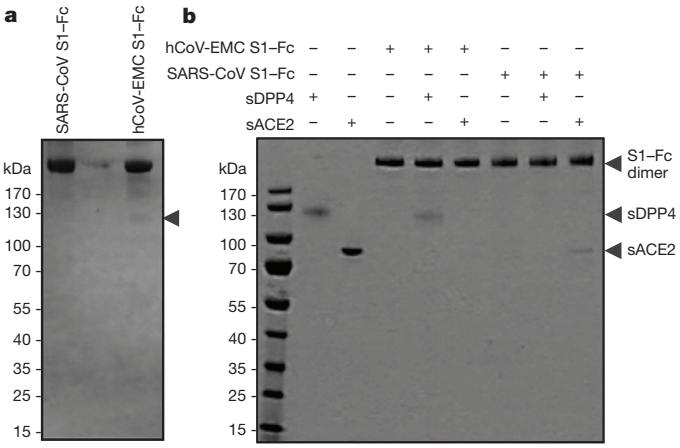


Figure 2 | hCoV-EMC S1 binding to DPP4. **a**, Huh-7 cell lysates were incubated with hCoV-EMC and SARS-CoV S1–Fc proteins and affinity-isolated proteins were subjected to protein electrophoresis under non-reducing conditions. The arrowhead indicates the position of the ~110-kDa DPP4 protein specifically isolated using the hCoV-EMC S1–Fc protein. **b**, hCoV-EMC and SARS-CoV S1–Fc proteins were mock-incubated or incubated with soluble DPP4 (sDPP4) or soluble ACE2 (sACE2) followed by protein A sepharose affinity isolation and subjected to protein electrophoresis under non-reducing conditions.

converting enzyme 2 (ACE2) for SARS-CoV[13]). In addition, sialic acid may act as a receptor for some coronaviruses[14].

Our initial experiments indicated that hCoV-EMC does not use ACE2 as an entry receptor[15]. Therefore we first examined whether the amino-terminal receptor-binding spike domain S1 binds to cells and investigated its correlation with cell susceptibility. We expressed the S1 domain fused to the Fc region of human IgG, yielding a recombinant disulphide-bonded dimer of approximately 280 kDa (Supplementary Fig. 1). Highly specific binding was observed to African green monkey kidney (Vero) and human liver (Huh-7) cells by immunofluorescence and fluorescence-activated cell sorting (FACS) analysis, whereas kidney cells of the *P. pipistrellus* bat showed intermediate staining (Fig. 1). No S1 binding was detectable to COS-7 African green monkey kidney cells (Fig. 1b). Furthermore, no specific binding to any of these cells was observed with a feline coronavirus S1 domain, whereas feline cells (*Felis catus* whole fetus, FCWF) showed strong reactivity (Supplementary Table 1). Binding of hCoV-EMC S1 was shown to correlate with susceptibility to hCoV-EMC infection and with viral genome detection in the culture medium of infected cells (Fig. 1). The hCoV-EMC S1 domain was demonstrated also to bind to cells from other species but its overall reactivity was more restricted compared to that observed for SARS-CoV S1 (Supplementary Table 1).

To identify the cell surface protein(s) binding to S1 we affinity-isolated proteins from Vero and Huh-7 cells using the S1–Fc chimaeric proteins. The hCoV-EMC S1–Fc protein—but not SARS-CoV S1–Fc—extracted a protein of ~110 kDa when analysed under non-reducing conditions from Huh-7 cell lysates (Fig. 2a). Mass spectrometric analysis identified this protein as dipeptidyl peptidase 4 (DPP4 or DPP IV, also called CD26; Supplementary Fig. 2). Similar results were obtained using Vero cells (data not shown). We subsequently produced soluble (that is, non-membrane-anchored) forms of DPP4 and ACE2 and found that the hCoV-EMC S1–Fc protein bound the former but not the latter, whereas the opposite was seen with the SARS S1–Fc protein (Fig. 2b). Soluble DPP4, but not soluble ACE2, inhibited infection of Vero cells by hCoV-EMC (Supplementary Fig. 3). Moreover, transient expression of human DPP4 in the non-susceptible COS-7 cells rendered these cells susceptible to binding the hCoV-EMC S1–Fc protein to their surface (Supplementary Fig. 4). These data indicate a direct and specific binding of hCoV-EMC S1 to human DPP4.

The DPP4 protein displays high amino acid sequence conservation across different species, including the sequence that we obtained from *P. pipistrellus* bat cells (Supplementary Fig. 5), particularly towards the carboxy-terminal end (Supplementary Fig. 6). Next, we tested surface

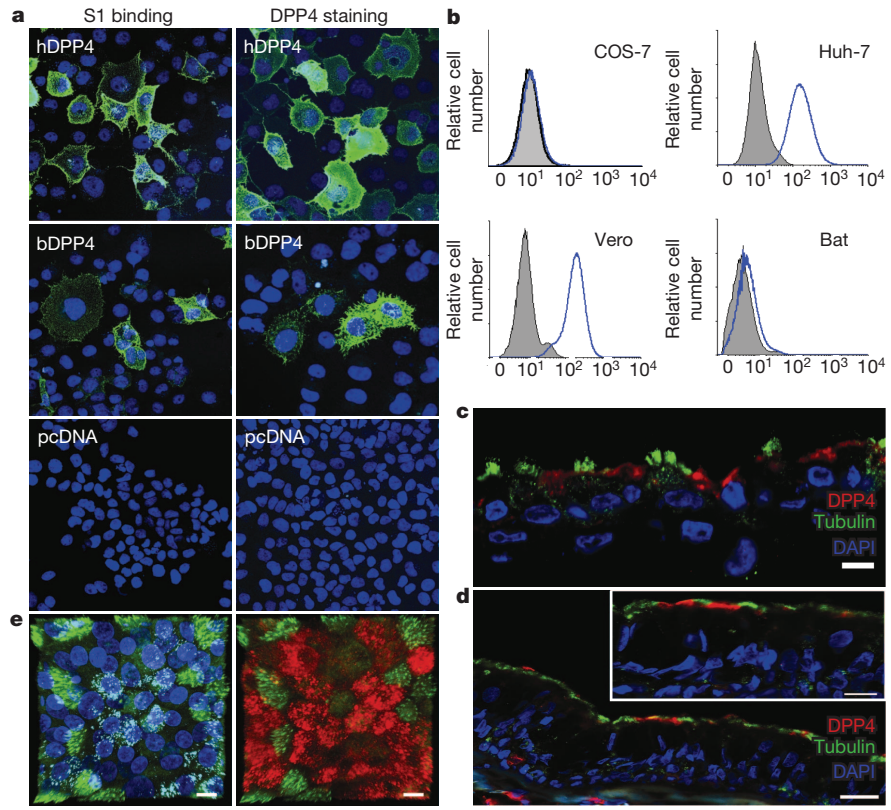


Figure 3 | DPP4 is present on hCoV-EMC-susceptible cell lines and human bronchiolar epithelial cells. **a**, COS-7 cells transfected with plasmids encoding human DPP4 (hDPP4), bat DPP4 (bDPP4) or a control plasmid (pcDNA) were tested for S1 binding and staining with a polyclonal antiserum against DPP4. **b**, Similarly, COS-7, Huh-7, Vero and bat cells were tested for reactivity with the same antiserum against DPP4 (blue lines) or with a control normal goat serum (grey peak). **c**, **d**, DPP4 expression (red) was also found in primary human bronchiolar epithelial cell cultures (**c**) and human bronchiolar tissue (**d**) and appeared to be localized to the apical surfaces of non-ciliated cells that do not express β-tubulin IV (green). **e**, Double-stranded viral RNA (cyan) was detected in hCoV-EMC-infected primary human bronchiolar epithelial cell cultures and appeared to be localized to non-ciliated cells that express DPP4 (red). Stainings were performed using antibodies directed against β-tubulin IV (ciliated cells; green), DPP4 (red), dsRNA (hCoV-EMC; cyan), and DAPI (cell nucleus; blue). All scale bars are 10 μm.

©2013 Macmillan Publishers Limited. All rights reserved

expression of DPP4 on susceptible and non-susceptible cells using a polyclonal human DPP4 antiserum. The specific reactivity of the anti-DPP4 serum with human and bat DPP4 was demonstrated by staining of human and bat DPP4-transfected cells (Fig. 3a). Expression of bat DPP4 in COS-7 cells allowed hCoV-EMC S1–Fc cell surface binding (Fig. 3a). Consistent with their susceptibility to hCoV-EMC infection and with the hCoV-EMC S1 cell surface binding, Vero and Huh-7 cells expressed high levels of DPP4 on their surface as judged by antibody reactivity, bat cells displayed low-level antibody binding, whereas COS-7 cells did not show any significant binding (Fig. 3b). Thus, DPP4 cell surface expression on the cell lines was consistent with hCoV-EMC S1 cell surface binding and with susceptibility to hCoV-EMC infection. The relevance of these observations was enforced by the finding that DPP4 expression was also found in primary human bronchiolar epithelial cell cultures (Fig. 3c) and in human bronchial lung tissue (Fig. 3d), in both instances localized to the apical surfaces of non-ciliated (tubulin-IV negative) cells. In addition, hCoV-EMC infection of human bronchiolar epithelial cell cultures appeared to be localized to the non-ciliated cells that express DPP4 (Fig. 3e).

To determine whether DPP4 essentially contributes to infection, susceptible cells were pre-incubated with polyclonal DPP4 antiserum before virus inoculation. Infection of Huh-7 cells was blocked by this serum but not by control serum or ACE2 antibodies, as evidenced by a strong reduction of virus infection, of virus excretion and of virus-induced cytopathic effects (Fig. 4a and Supplementary Fig. 7). In addition, infection of primary bronchiolar epithelial cells was blocked by the DPP4 antibodies in a dose-dependent manner (Fig. 4b). We next examined whether DPP4 expression confers susceptibility to hCoV-EMC infection. COS-7 cells transfected with the human DPP4 expression plasmid, but not with the empty plasmid, were efficiently infected by hCoV-EMC as demonstrated by the presence of viral non-structural proteins in the cells (Fig. 4c) and of viral RNA (Fig. 4d) and infectious virus in the cell supernatants (not shown). Likewise, expression of bat DPP4 in COS-7 cells conferred susceptibility to the virus, although to a lesser extent (Fig. 4c). Transfection of human DPP4 in non-susceptible cells of different species origin (feline, murine and canine) also permitted infection with hCoV-EMC (Supplementary Fig. 8a), whereas other human coronaviruses such as hCoV-NL63, hCoV-229E and hCoV-OC43 were not able to infect human DPP4-transfected cells (Supplementary Fig. 8b–d). Collectively, our data demonstrate that DPP4 of human and bat origin acts as a functional receptor for hCoV-EMC.

After ACE2 and APN, DPP4 is the third exopeptidase to be discovered as a receptor for coronaviruses. DPP4 is a multifunctional 766-amino-acid-long type-II transmembrane glycoprotein presented in a dimeric form on the cell surface. It preferentially cleaves dipeptides from hormones and chemokines after a proline amino acid residue, thereby regulating their bioactivity[16]. Yet the use of peptidases by coronaviruses may be more related to their abundant presence on epithelial and endothelial tissues—the primary tissues of coronavirus infection—rather than to their proteolytic activity, which in the case of APN and ACE2 appeared not to be critical for infection[12,13]. Consistently, hCoV-EMC infection could not be blocked by the DPP4 inhibitors sitagliptin, vildagliptin, saxagliptin or P32/98 (Supplementary Fig. 9). DPP4 also has a major role in glucose metabolism by its degradation of incretins and has been further implicated in T-cell activation, chemotaxis modulation, cell adhesion, apoptosis and regulation of tumorigenicity[16,17]. In humans it is primarily expressed on the epithelial cells in kidney, small intestine, liver and prostate, and on activated leukocytes, whereas it occurs as a soluble form in the circulation[16,17]. At present little is known about the tropism of hCoV-EMC *in vivo*; the virus has been detected only in upper respiratory swabs, urine, sputum and tracheal aspirate[3,4]. Our observation of DPP4 being expressed on non-ciliated bronchial epithelial cells together with its reported expression in the kidney is consistent with clinical manifestations of hCoV-EMC infection. It is important to note that most

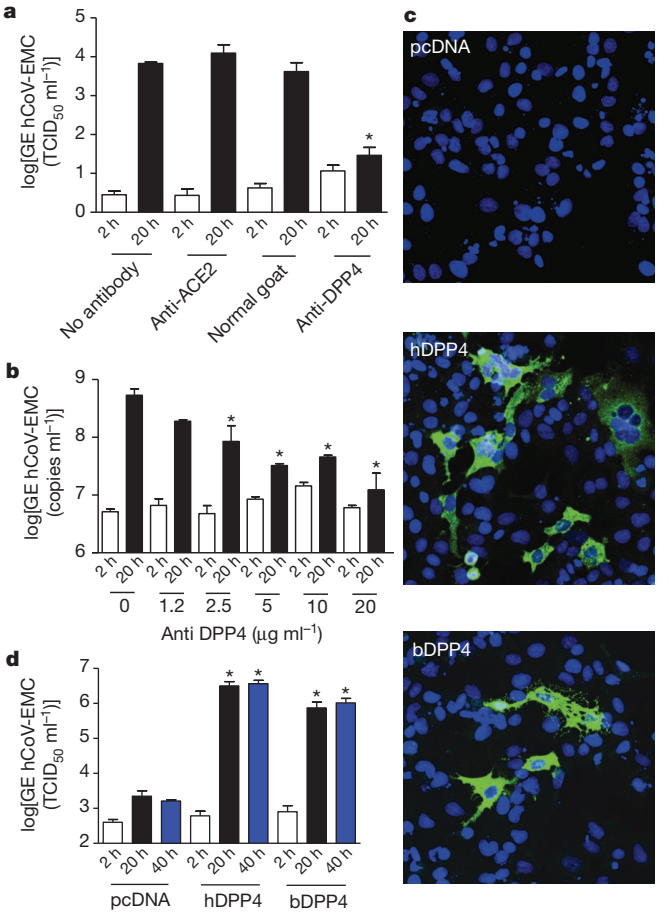


Figure 4 | DPP4 is essential for virus infection. **a**, Inhibition of hCoV-EMC infection of Huh-7 cells by antibodies to DPP4. Supernatants collected at 2 h (open bars) and 20 h (filled bars) were tested for the presence of hCoV-EMC RNA using a TaqMan assay. Results representative of three different experiments are shown as ΔCt values (one-way ANOVA test, $*P < 0.05$; $n = 3$ per group), normal goat, normal goat serum. **b**, Infection of human primary bronchiolar epithelial cells is blocked by the DPP4 antibodies in a dose-dependent manner and samples were analysed at 2 h (open bars) and 20 h (filled bars) after infection (one-way ANOVA test, $*P < 0.05$; $n = 3$ per group). **c**, COS-7 cells transfected with plasmids encoding human DPP4 (hDPP4), bat DPP4 (bDPP4) or a control plasmid (pcDNA) were inoculated with hCoV-EMC at a multiplicity of infection of 1 and left for 1 h. Cells were washed twice and stained at 8 h after infection (original magnification, ×200) or supernatant collected at 2 h (open bars), 20 h (black bars) and 40 h (blue bars) was tested for the presence of hCoV-EMC RNA using a TaqMan assay (**d**). Results representative of three different experiments are expressed as GE ($TCID_{50}$ ml^{-1}) values (one-way ANOVA test, $*P < 0.05$; $n = 4$ per group). All error bars represent s.e.m.

respiratory viruses, including SARS-CoV[18], have a marked tropism for ciliated cells that are more widely distributed along the upper and lower respiratory tract.

The epidemiological history of hCoV-EMC remains enigmatic. As for SARS-CoV and hCoV-NL63 (refs 7, 19), a bat origin possibly combined with the existence of an intermediate animal reservoir seems feasible. In view of the evolutionary conservation of DPP4 and the ability of hCoV-EMC to use bat DPP4 as a functional receptor, such host species switching would not be surprising. Adaptive mutations in the SARS-CoV S1 domain allowing improved binding to human ACE2 have been noted, explaining, at least in part, the zoonotic transmission event[20,21]. Further in-depth characterization of the binding interface of hCoV-EMC S1 and DPP4 may shed light on the possible adaptive processes of this virus or related coronaviruses using DPP4 in novel host species.

©2013 Macmillan Publishers Limited. All rights reserved

Variations in soluble DPP4 levels in serum have been reported as clinically relevant in a number of pathophysiological conditions including type-2 diabetes mellitus and virus infections[16,17]. It will be important to investigate whether and how varying soluble DPP4 levels affect hCoV-EMC pathogenesis. Downregulation of ACE2 expression after SARS-CoV infection has been shown to contribute to the severity of disease[22], consistent with the protective role of soluble ACE2 in lung injury[23]. Given the importance of DPP4 in regulation of chemokine and cytokine responses[16] one may speculate that *in vivo* downregulation of the hCoV-EMC receptor may potentially influence the pathogenesis of this virus. Preliminary findings *in vitro*, however, indicate that S1 binding to DPP4 did not result in significant downregulation of DPP4 or of the DPP4 enzymatic activity on Huh-7 cells (not shown), possibly due to the observed active recycling of DPP4 from the plasma membrane[16]. Manipulation of DPP4 levels or development of inhibitors that target the binding interface between the S1 domain and receptor *in vivo* may provide therapeutic opportunities to combat hCoV-EMC infection. In addition, future studies should address the development of effective vaccines, including those that elicit antibodies that prevent the binding of hCoV-EMC to DPP4.

METHODS SUMMARY

hCoV-EMC infection and detection. Virus stocks of hCoV-EMC were prepared as described earlier[5]. Vero, COS-7, Huh-7 and kidney cells of the *P. pipistrellus* bat[15] were inoculated with hCoV-EMC for 1 h and incubated with medium containing 1% fetal bovine serum. Formaldehyde-fixed cells were stained using rabbit-anti-SARS-CoV NSP4 antibodies that are cross-reactive for hCoV-EMC, according to standard protocols using a FITC-conjugated swine-anti-rabbit antibody as a second step. Viral replication was quantified by quantitative PCR as described in Methods.

Protein expression. S1 receptor-binding domains of hCoV-EMC, SARS-CoV and FIPV fused to the Fc region of human IgG, and soluble versions of DPP4 and ACE2 were expressed and purified as described in the Methods.

DPP4 cell surface expression measurement using S1–Fc proteins. S1 binding of cells was measured by incubating 2.5×10^5 cells with 15 μg ml^{-1} of S1–Fc followed by incubation with FITC or DyLight-488-labelled goat-anti-human IgG antibody and analysis by flow cytometry.

Full Methods and any associated references are available in the online version of the paper.

Received 3 December 2012; accepted 13 February 2013.

1. Weiss, S. R. & Navas-Martin, S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol. Mol. Biol. Rev.* **69,** 635–664 (2005).
2. Peiris, J. S. M., Guan, Y. & Yuen, K. Y. Severe acute respiratory syndrome. *Nature Med.* **10,** S88–S97 (2004).
3. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367,** 1814–1820 (2012).
4. Bermingham, A. *et al.* Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro Surveill.* **17,** 20290 (2012).
5. van Boheemen, S. *et al.* Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio* **3,** e00473–12 (2012).
6. Graham, R. L. & Baric, R. S. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J. Virol.* **84,** 3134–3146 (2010).
7. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310,** 676–679 (2005).
8. World Health Organisation. Interim surveillance recommendations for human infection with novel coronavirus. http://www.who.int/csr/don/2013_02_21/en/index.html (21 February 2013).
9. Drexler, J. F. *et al.* Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J. Virol.* **84,** 11336–11349 (2010).
10. Williams, R. K., Jiang, G. S. & Holmes, K. V. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proc. Natl Acad. Sci. USA* **88,** 5533–5536 (1991).
11. Yeager, C. L. *et al.* Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357,** 420–422 (1992).
12. Delmas, B. *et al.* Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature* **357,** 417–420 (1992).
13. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426,** 450–454 (2003).
14. Schultze, B. & Herrler, G. Bovine coronavirus uses *N*-acetyl-9-*O*-acetylneuraminic acid as a receptor determinant to initiate the infection of cultured cells. *J. Gen. Virol.* **73,** 901–906 (1992).
15. Müller, M. A. *et al.* Human coronavirus EMC does not require the SARS-coronavirus receptor and maintains broad replicative capability in mammalian cell lines. *mBio* **3,** e00515–12 (2012).
16. Lambeir, A. M., Durinx, C., Scharpe, S. & De Meester, I. Dipeptidyl-peptidase IV from bench to bedside: an update on structural properties, functions, and clinical aspects of the enzyme DPP IV. *Crit. Rev. Clin. Lab. Sci.* **40,** 209–294 (2003).
17. Boonacker, E. & Van Noorden, C. J. The multifunctional or moonlighting protein CD26/DPPIV. *Eur. J. Cell Biol.* **82,** 53–73 (2003).
18. Sims, A. C., Burkett, S. E., Yount, B. & Pickles, R. J. SARS-CoV replication and pathogenesis in an *in vitro* model of the human conducting airway epithelium. *Virus Res.* **133,** 33–44 (2008).
19. Huynh, J. *et al.* Evidence supporting a zoonotic origin of human coronavirus strain NL63. *J. Virol.* **86,** 12816–12825 (2012).
20. Wu, K., Peng, G., Wilken, M., Geraghty, R. J. & Li, F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J. Biol. Chem.* **287,** 8904–8911 (2012).
21. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309,** 1864–1868 (2005).
22. Kuba, K. *et al.* A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nature Med.* **11,** 875–879 (2005).
23. Imai, Y. *et al.* Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* **436,** 112–116 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Snijder for providing the anti-SARS-CoV NSP4 antibody, and E. Kindler, H. Jónsdóttir, R. Rodriguez, T. Bestebroer, S. Pas, G. Arron, M. van Velzen and W. Ouwendijk for technical assistance. This work was supported by a fellowship from China Scholarship Council to H.M. The study was financed by the European Union FP7 project EMPERIE (contract number 223498), ANTIGONE (contract number 278976), the Swiss National Science Foundation (31003A_132898) and the 3R Research Foundation Switzerland (Project 128-11). C.D. was supported by the German Research Foundation (DFG grant DR 772/3-1, KA1241/18-1) and the German Ministry of Education and Research (BMBF SARS II).

Author Contributions B.J.B., V.S.R. and B.L.H. designed and coordinated the study. B.J.B. and B.L.H. contributed equally to the study. V.S.R., H.M., S.L.S., D.H.W.D., M.A.M., R.D., D.M., B.J.B. and B.L.H conducted the experiments. A.Z. provided the virus, and R.A.M.F., J.A.A.D., V.T., C.D., A.D.M.E.O. and P.J.M.R. supervised part of the experiments. All authors contributed to the interpretations and conclusions presented. B.J.B. and B.L.H. wrote the manuscript, and A.D.M.E.O. and P.J.M.R. participated in editing it.

Author Information The *P. pipistrellus* bat *DPP4* sequence has been deposited in GenBank/EMBL/DDBJ under accession number KC249974. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.J.B. (b.j.bosch@uu.nl) or B.L.H. (b.haagmans@erasmusmc.nl).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Protein expression. A plasmid encoding hCoV-EMC S1–Fc was generated by ligating a fragment encoding the S1 domain (residues 1–747) 3′ terminally to a fragment encoding the Fc domain of human IgG into the pCAGGS expression vector. Likewise, an S1–Fc expression plasmid was made for the SARS-CoV domain S1 subunit (isolate CUHK-W1; residues 1–676), the FIPV S1 domain (isolate 79–1146; residues 1–788) and the human ACE2 ectodomain (sACE2; residues 1–614). Fc chimaeric proteins were expressed by transfection of the expression plasmids into HEK-293T cells and affinity purified from the culture supernatant using protein A sepharose beads (GE Healthcare). Purified ACE2–Fc was cleaved with thrombin and soluble ACE2 was purified by gel-filtration chromatography. A DPP4 expression vector was generated by cloning the full-length human *DPP4* gene into the pCAGGS or the pcDNA3 vector. A plasmid encoding the ectodomain of human DPP4 was generated by ligating a fragment encoding residues 39–766 of human *DPP4* into a pCD5 expression vector[24] encoding the signal sequence of CD5 and an OneSTrEP affinity tag (IBA GmbH). Soluble DPP4 ectodomain was prepared by transfection of the expression plasmid into HEK-293T cells and affinity purification from the culture supernatant using Strep-Tactin sepharose beads (IBA GmbH). S1 binding of cells was measured by incubating 2.5×10^5 cells with 15 μg ml^{-1} of S1–Fc followed by incubation with FITC or DyLight-488-labelled goat-anti-human IgG antibody and analysis by flow cytometry.

Affinity-purification and detection of DPP4. Huh-7 cells were washed twice with ice-cold PBS, scraped off the plastic with a rubber policeman, pelleted and lysed in ice-cold lysis buffer (0.3% n-decyl-β-D-maltopyranoside in PBS) containing protease inhibitors (Roche Complete Mini and phenylmethylsulphonyl fluoride) at a final density of ~2.5×10^7 cells ml^{-1}. The supernatants of centrifuged cell lysates were precleared with protein A sepharose beads after which 10 μg of S1–Fc and 100 μl protein A sepharose beads (50% v/v) was added to 1 ml of cell lysate and incubated for 1 h at 4 °C under rotation. Beads were washed thrice with lysis buffer and once with PBS and subjected to protein electrophoresis (NoVEX 4–12% Tris-glycine gradient gel, Invitrogen) under non-reducing conditions. A distinct ~110-kDa protein co-purified with hCoV-EMC S1–Fc was visualized by GelCodeBlue staining, excised from the gel, incubated with trypsin and analysed by mass spectrometry.

Mass spectrometry and data analysis. The distinct ~110-kDa protein which co-purified with hCoV-EMC S1–Fc was excised from the gel and subjected to in-gel reduction with dithiothreitol, alkylation with chloroacetamide and digestion with trypsin (Promega, sequencing grade), essentially as described[25]. Alternatively, affinity-isolated proteins were reduced and alkylated on beads similarly as described above. Nanoflow LC-MS/MS was performed on either an 1100 series capillary LC system (Agilent Technologies) coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo), or an EASY-nLC coupled to a Q Exactive mass spectrometer (Thermo), operating in positive mode and equipped with a nanospray source. Peptide mixtures were trapped on a ReproSil C18 reversed-phase column (Dr Maisch GmbH; column dimensions 1.5 cm × 100 μm, packed in-house) at a flow rate of 8 μl min^{-1}. Peptide separation was performed on ReproSil C18 reversed-phase column (Dr Maisch GmbH; column dimensions 15 cm × 50 μm, packed in-house) using a linear gradient from 0% to 80% B (A = 0.1% formic acid; B = 80% (v/v) acetonitrile, 0.1% formic acid) in 70 or 120 min and at a constant flow rate of 200 nl min^{-1}. The column eluent was directly sprayed into the ESI source of the mass spectrometer. Mass spectra were acquired in continuum mode; fragmentation of the peptides was performed in data-dependent mode by CID or HCD. Peak lists were automatically created from raw data files using the Mascot Distiller software (version 2.3; MatrixScience) or Proteome Discoverer (version 1.3; Thermo). The Mascot algorithm (version 2.2; MatrixScience) was used for searching against a Uniprot database (release 2012_10.fasta, taxonomy: *Homo sapiens* or *Chlorocebus sabaeus*). The peptide tolerance was set to 10 p.p.m. and the fragment ion tolerance was set to 0.8 Da for CID spectra (LTQ-Orbitrap) or to 20 milli unified atomic mass unit (mu) for HCD (Q Exactive) spectra. A maximum number of two missed cleavages by trypsin were allowed and carbamidomethylated cysteine and oxidized methionine were set as fixed and variable modifications, respectively.

Blocking of hCoV-EMC replication by DPP4 antiserum. Huh-7 and primary airway epithelial cells (triplicates of one donor) were pre-incubated with antibodies to DPP4 (polyclonal goat-anti DPP4 immunoglobulin, R&D systems) in a range of 0–20 μg ml^{-1}. Cells were infected with hCoV-EMC, incubated for 1 h at 37 °C, washed and subsequently incubated with medium containing the respective antibody concentrations. Supernatants were collected at 2 h and 20 h and were analysed for the presence of hCoV-EMC RNA using a real-time Taqman assay.

Human bronchial epithelial cultures and confocal microscopy analysis. Human bronchial lung tissue was obtained from patients (age >18 years old) who underwent surgical lung resection in their diagnostic pathway for any pulmonary disease and that gave informed consent. This was done in accordance with local regulation of the Kantonal Hospital St Gallen, Switzerland, as part of the St Gallen Lung Biopsy Biobank (SGLBB) which received approval by the ethics committee of the Kanton St Gallen (EKSG 11/044, 27 April 2011; and EKSG 11/103, 23 September 2011).

Primary human bronchial epithelial cultures were generated as previously described[26]. Human bronchial epithelial cultures were maintained for 1–2 months until pseudostratified and fully differentiated epithelia were obtained. Human bronchial tissue or human bronchial epithelial cultures were fixed with 4% PFA (FormaFix) for 30 min at room temperature. The fixed human bronchial tissue or cultures were mounted in Tissue-Tek OCT medium and snap frozen in liquid nitrogen, from which 10 μm horizontal sections were made. The cryosections were immunostained using the procedure as described[26], using mouse monoclonal antibody anti-β-tubulin IV (Sigma) and goat anti-hDPP4 polyclonal antibody (R&D Systems) as primary antibodies, and Dylight-488-labelled, anti-mouse IgG (H+L) and Dylight-549-labelled, anti-goat IgG (H+L) as secondary antibodies (Jackson ImmunoResearch). Counterstaining was done with DAPI (Invitrogen). Fluorescent images were acquired using Plan Apochromat 63×/1.40 oil DIC M27 or EC Plan-Neofluor 40×/1.30 oil DIC M27 objectives on a Zeiss LSM 710 confocal microscope. Image capture, analysis and processing were performed using the ZEN 2010 (Zeiss) software packages.

hCoV-EMC-infected human bronchial epithelial cultures were fixed with 4% PFA (FormaFix) for 30 min at room temperature. Fixed cultures were immunostained using the procedure as described[26], with mouse monoclonal antibody anti-dsRNA (J2, English and Scientific Consulting Bt.) and goat anti-hDPP4 polyclonal antibody (R&D Systems) as primary antibodies, and Dylight-488-labelled, anti-mouse IgG (H+L) and Dylight-549-labelled, anti-goat IgG (H+L) as secondary antibodies (Jackson ImmunoResearch). The rabbit monoclonal anti-β-tubulin conjugated with Alexafluor 647 (9F3, Cell Signaling) was applied as tertiary antibody. Counterstaining was done with DAPI (Invitrogen). *Z*-stack images were acquired using a Plan Apochromat 63×/1.40 oil DIC M27 objective on a Zeiss LSM 710 confocal microscope from which a three-dimensional reconstruction was made using maximum intensity projection. Image capture, analysis and processing were performed using the ZEN 2010 (Zeiss) and Imaris (Bitplane Scientific Software) software packages.

DPP4 enzymatic activity. DPP4 activity was measured on live cells using DPPIV/CD26 Assay kit for Biological Samples (Enzo Life sciences). Briefly, Vero cells (1×10^4 cells well^{-1}) were grown in 96-well plates, washed twice with PBS, incubated with or without inhibitor (20 μg ml^{-1}) for 10 min after which H-Gly-Pro-AMC substrate was added and incubated for 10 min. Fluorescence intensity was measured at 380/460 nm using Tecan Infinite F200.

Cloning and expression of human and bat DPP4. Total RNA was isolated from Huh-7 and PipNi/1 cells using RNeasy mini kit (Qiagen) and cDNAs were synthesized by using the Superscript reverse transcriptase (Life Technologies). The complete *DPP4* genes were amplified using Pfu Ultra II fusion HS DNA polymerase (Stratagene) and cloned into the pcDNA 3.1 (+) expression vector (Life Technologies). After sequencing, alignment was performed using ClustalW in the MEGA 5.0 software package (http://www.megasoftware.net) and the trees were constructed by using the neighbour-joining method with *p*-distance (gap/missing data treatment; complete deletion) and 1,000 bootstrap replicates as in MEGA 5.0. pcDNA plasmids containing the human or bat DPP4 or the empty pcDNA plasmid were transfected into COS-7 cells using lipofectamine 2000 (Life technologies). After 24 h incubation, cells were stained with both goat-anti-human DPP4 polyclonal antibody (R&D system) and a rabbit anti-goat IgG–FITC antibody (Sigma).

Virus infection, RNA extraction and quantitative RT–PCR. Virus stocks of hCoV-EMC were prepared as described previously[5]. Vero, COS-7, Huh-7 and kidney cells of the *P. pipistrellus* bat cells[15] were inoculated with hCoV-EMC for 1 h and incubated with medium containing 1% fetal bovine serum. Formaldehyde-fixed cells were stained using rabbit-anti-SARS-CoV NSP4 antibodies that are cross-reactive for hCoV-EMC, according to standard protocols using a FITC-conjugated swine-anti-rabbit antibody as a second step. RNA from 200 μl of culture supernatant was isolated with the Magnapure LC total nucleic acid isolation kit (Roche) and eluted in 100 μl. hCoV-EMC RNA was quantified on the ABI prism 7700, with the TaqMan Fast Virus 1-Step Master Mix (Applied Biosystems) using 20 μl isolated RNA, 1× Taqman mix, 0.5 U uracil-*N*-glycosylase, 45 pmol forward primer (5′-GGGTGTACCTCTTAATGCCAATTC-3′), 45 pmol reverse primer (5′-TCTGTCCTGTCTCCGCCAAT-3′) and 5 pmol probe (5′-FAM-ACCCCTGCGCAAAATGCTGGG-BHQ1-3′) . Amplification parameters were 5 min at 50 °C, 20 s at 95 °C, and 45 cycles of 3 s at 95 °C, and 30 s at 60 °C. RNA dilutions isolated from an hCoV-EMC stock were used as a standard. hCoV-NL63, hCoV-229E and hCoV-OC43 quantitative PCRs were routinely performed at the diagnostics Department of Viroscience at the Erasmus MC Rotterdam.

©2013 Macmillan Publishers Limited. All rights reserved

Statistics. We compared the mean Ct values and log GE hCoV-EMC using one-way ANOVA with post-test Bonferroni. Statistical analysis was performed with Prism 4.0 (Graphpad).

24. Bosch, B. J. *et al.* Recombinant soluble, multimeric HA and NA exhibit distinctive types of protection against pandemic swine-origin 2009 A(H1N1) influenza virus infection in ferrets. *J. Virol.* **84,** 10366–10374 (2010).
25. van den Berg, D. L. *et al.* An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6,** 369–381 (2010).
26. Dijkman, R. *et al.* Isolation and characterization of current human coronavirus strains in primary human epithelia cultures reveals differences in target cell tropism. *J. Virol.* http://dx.doi.org/10.1128/JVI.03368-12 (2013).

©2013 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature11941

USP33 regulates centrosome biogenesis via deubiquitination of the centriolar protein CP110

Ji Li[1], Vincenzo D'Angiolella[1]†, E. Scott Seeley[2], Sehyun Kim[1], Tetsuo Kobayashi[1], Wenxiang Fu[1], Eric I. Campos[3], Michele Pagano[1,4] & Brian David Dynlacht[1]

Centrosome duplication is critical for cell division, and genome instability can result if duplication is not restricted to a single round per cell cycle. Centrosome duplication is controlled in part by CP110, a centriolar protein that positively regulates centriole duplication while restricting centriole elongation and ciliogenesis. Maintenance of normal CP110 levels is essential, as excessive CP110 drives centrosome over-duplication and suppresses ciliogenesis, whereas its depletion inhibits centriole amplification and leads to highly elongated centrioles and aberrant assembly of cilia in growing cells[1,2]. CP110 levels are tightly controlled, partly through ubiquitination by the ubiquitin ligase complex $SCF^{cyclin\ F}$ during G2 and M phases of the cell cycle[3]. Here, using human cells, we report a new mechanism for the regulation of centrosome duplication that requires USP33, a deubiquitinating enzyme that is able to regulate CP110 levels. USP33 interacts with CP110 and localizes to centrioles primarily in S and G2/M phases, the periods during which centrioles duplicate and elongate. USP33 potently and specifically deubiquitinates CP110, but not other cyclin-F substrates. USP33 activity antagonizes $SCF^{cyclin\ F}$-mediated ubiquitination and promotes the generation of supernumerary centriolar foci, whereas ablation of USP33 destabilizes CP110 and thereby inhibits centrosome amplification and mitotic defects. To our knowledge, we have identified the first centriolar deubiquitinating enzyme whose expression regulates centrosome homeostasis by countering cyclin-F-mediated destruction of a key substrate. Our results point towards potential therapeutic strategies for inhibiting tumorigenesis associated with centrosome amplification.

We recently characterized a novel human CP110-interacting protein, NEURL4, a daughter-centriole-specific protein that inhibits the formation of ectopic microtubule-organizing centres[4]. During the course of these studies, NEURL4 was independently found to interact with USP20 (also known as VDU2) in a proteome-wide screen for deubiquitinating enzyme (DUB)-associated proteins[5]. USP20 is highly related to a second DUB, USP33 (also known as VDU1), with which it shares ~59% amino acid identity, explaining their overlapping functions[6,7]. Thus, we sought to examine whether USP20 and USP33 interact with the CP110–NEURL4 complex to regulate centrosome function. We immunoprecipitated each protein from human 293T-cell lysates, and found that endogenous USP33 and USP20 specifically interact with CP110 and NEURL4, but not with two other CP110-interacting proteins, CEP97 and cyclin F (Fig. 1a). Moreover, these interactions were also observed in cells ectopically expressing Flag-tagged CP110 (Flag–CP110) or Flag–USP33 (Fig. 1b and Supplementary Fig. 1a–e). The interactions with USP20 and USP33 are probably direct because (1) bacterially expressed glutathione *S*-transferase (GST)–USP33 and USP20 bind to *in vitro* translated CP110 in binding assays (Supplementary Fig. 1f), and (2) NEURL4 depletion does not affect the binding between USP33 and CP110 (Supplementary Fig. 1g). We found that endogenous USP33 and CP110 co-fractionate in at least three distinct protein complexes ranging from 0.4–1.3 megadaltons (MDa). This suggests the existence of larger CP110–USP33 deubiquitinase complexes that include proteins other than NEURL4 (Supplementary Fig. 2). Both USP33 and USP20 share three functional domains, a zinc-finger ubiquitin-binding motif (ZnF-UBP), a catalytic domain, and two domains present in ubiquitin-specific proteases (DUSP). We investigated the interaction between USP33 and CP110, and found that CP110 binds to a region of USP33 within its catalytic domain (Fig. 1b).

USP33 and USP20 have been shown to partition to the cytoplasm, including the endoplasmic reticulum[8], and a short isoform of USP33 localizes to the Golgi apparatus[9]. Because CP110 is a centrosomal protein, we visualized endogenous USP33, USP20, CP110 and other centrosomal markers in human osteosarcoma (U2OS) cells by indirect immunofluorescence. We observed USP33 localization to the cytoplasm, as expected, and we also found that USP33 co-localized with centrin 2 and γ-tubulin, which are centrosomal markers (Fig. 2a, b and Supplementary Fig. 3a, b). Notably, USP33 partially co-localized with CP110 at centrioles. USP33 localization appeared to be more proximal than CP110, suggesting that the former protein interacts with a subpopulation of CP110. A similar localization pattern was observed for

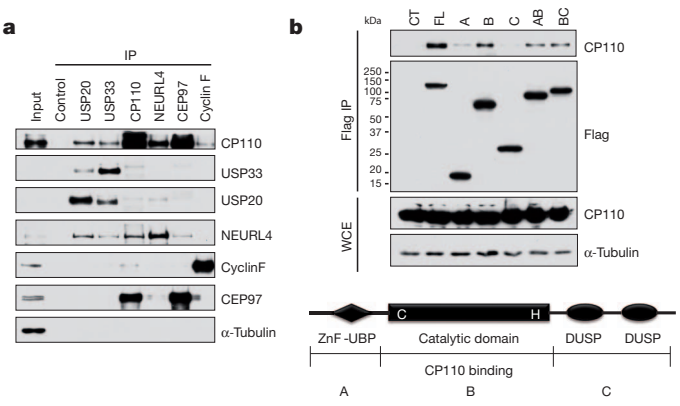


Figure 1 | Identification of USP33 as a CP110-associated protein. **a**, HEK293T cell extracts were immunoprecipitated (IP) with the indicated antibodies (top), and western blots were probed as shown (right). Non-specific rabbit IgG was used as control. α-Tubulin is shown as a negative control. **b**, Bottom, domains of USP33 are indicated. The catalytic domain of USP33 stably binds to CP110. Double cysteine and histidine point mutations in the catalytic domain (DM mutant) render the DUB inactive. Top, 293T cells were transfected with haemagglutinin (HA)-tagged CP110 and Flag-tagged full-length or truncated fragments of USP33, extracts were immunoprecipitated, and western blots of the resulting immunoprecipitates were probed with anti-Flag and CP110 antibodies. CT, control, transfected with empty vector; FL, full-length USP33 protein. WCE, whole-cell extract.

[1]Department of Pathology and Cancer Institute, Smilow Research Center, New York University School of Medicine, 522 1st Avenue, New York, New York 10016, USA. [2]Department of Pathology, University of California San Francisco, San Francisco, California 94143, USA. [3]Department of Biochemistry, Smilow Research Center, New York University School of Medicine, 522 1st Avenue, New York, New York 10016, USA. [4]Howard Hughes Medical Institute, Department of Pathology and Cancer Institute, Smilow Research Center, New York University School of Medicine, 522 1st Avenue, New York, New York 10016, USA. †Present address: Gray Institute for Radiation Oncology and Biology, Department of Oncology, University of Oxford, Oxford OX3 7DQ, UK.

©2013 Macmillan Publishers Limited. All rights reserved

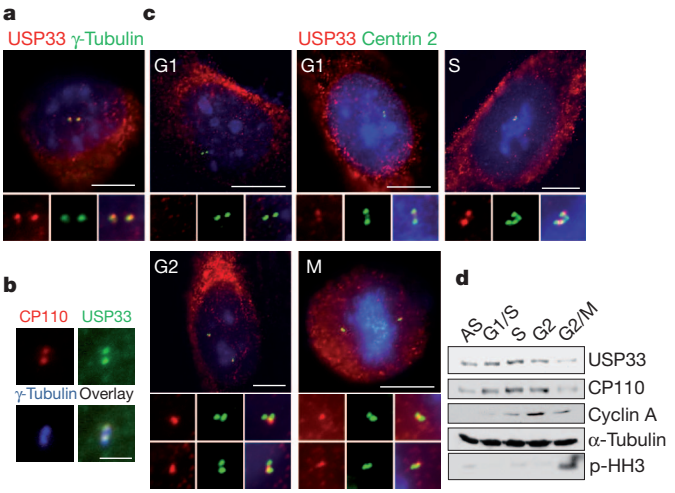


Figure 2 | USP33 localizes to centrosomes primarily during S and G2/M phases of the cell cycle. **a**, **b**, Co-localization of USP33 with CP110 and other centrosome and centriolar markers was visualized by immunofluorescence, as indicated. **c**, USP33 (red) and centrin 2 (green) were visualized throughout the cell cycle, as in panels **a**, **b**. DNA was stained with 4′,6-diamidino-2-phenylindole (DAPI; blue). U2OS cells were treated with mimosine for G1-phase cells or were subject to double-thymidine block and release to enrich for S- and G2-phase cells. Cells with two centrin-2 dots were counted as G1-phase cells; those with three to four clustered centrin-2 dots as S-phase cells; and those with three to four centrin-2 dots, in which two pairs of dots were at least 3 μm apart, were counted as G2-phase cells. Localization of USP33 to centrosomes was most prominent in S and G2/M phases. **d**, USP33 protein levels parallel those of CP110 during the cell cycle. Cells were synchronized by double-thymidine block and release and synchronization confirmed by fluorescence-activated cell sorting (FACS) analysis (see Supplementary Fig. 3e). AS, asynchronous; p-HH3, phospho-Ser-10-histone H3. Scale bars in **a**, **c**: 10 μm; **b**: 2 μm.

USP20 (data not shown). The centrosomal localization of USP33 is specific because it disappeared after USP33 was silenced with short interfering RNAs (siRNAs) (Supplementary Fig. 3c). Because CP110 levels are high during S phase and low during G2/M owing to cyclin-F-mediated ubiquitination, we sought to determine the levels of USP33 in synchronized U2OS cells. USP33 associated with centrosomes predominantly in S and G2 phases of the cell cycle but less prominently in G1 phase, mirroring the abundance of CP110 protein during the cell cycle (Fig. 2c and Supplementary Fig. 3d)[1,3]. Notably, USP33 protein levels also oscillated in parallel during cell-cycle progression, showing increased expression in S and early G2 phases and lower levels in late G2 and M phases (Figs 2d and 3b and Supplementary Fig. 3e). We found that centrosomal targeting of USP33 is partially dependent on CP110, because depletion of CP110 reduced USP33 localization to centrosomes (see later) (Supplementary Fig. 3f).

DUBs are able to remove ubiquitin from proteins and regulate substrate abundance or function[5]. As USP33 and USP20 interact with CP110 at centrosomes, and CP110 protein levels correlated with USP33 and USP20 during the cell cycle, we investigated whether USP33 or USP20 might stabilize CP110. Interestingly, depletion of USP33 with siRNAs reduced CP110 levels by ~50%, whereas depletion of USP20 had a very modest impact on CP110 levels (Fig. 3a), and co-depletion of USP20 with USP33 did not accentuate the impact of USP33 silencing. These results indicate that USP33 may have a more prominent role than USP20 in regulating CP110 stability; thus, we focused on the functional role of USP33. Importantly, cells treated with two different siRNAs targeting USP33 showed reduced levels of CP110, but not other CP110-interacting proteins, including NEURL4 and cyclin F (Supplementary Fig. 4a). Moreover, USP33 depletion led to decreased CP110 intensity at centrioles in both U2OS and human telomerase reverse transcriptase (hTERT)-immortalized retinal pigment epithelial cells (Supplementary Fig. 4b, c). USP33 also stabilized CP110 in cells enriched in S and G2 phases (Fig. 3b and Supplementary Fig. 4e, f). Furthermore, USP33 suppression significantly reduced CP110 protein half-life from ~4.5 h to ~2 h (Supplementary Fig. 4g, h). Because CP110 is a substrate of the $SCF^{cyclin\ F}$ ubiquitin ligase complex[3], we examined the impact of USP33 on another $SCF^{cyclin\ F}$ substrate, RRM2 (ref. 10). We observed that USP33 depletion did not lead to a reduction in RRM2 levels (Supplementary Fig. 4e, f), indicating that USP33 specifically stabilizes CP110.

Next, we examined whether stabilization of CP110 by USP33 is mediated through deubiquitination, because USP33 acts as a DUB. Remarkably, USP33 overexpression diminished CP110 ubiquitination provoked by ectopic production of cyclin F in cells (Fig. 3c). CP110 deubiquitination was markedly attenuated when a catalytically inactive USP33 mutant (USP33 DM), in which cysteine 163 in the active site was replaced with a serine (C163S) and histidine 683 was modified to glutamine (H642Q), was transfected[11]. In addition, overexpression of two unrelated DUBs, USP1 and USP37, did not promote CP110 deubiquitination (Fig. 3c and data not shown). To test whether deubiquitination by USP33 requires CP110 binding, first we mapped the region of CP110 that binds USP33 and found that the binding domain was located within the first 200 amino-terminal residues (Supplementary Fig. 1d). Notably, when we tested a truncation mutant unable to bind USP33 (CP110 ΔN), we found that USP33 was unable to deubiquitinate this substrate (Fig. 3d). In addition, we tested the activity of USP33 in an *in vitro* deubiquitination assay. We purified Flag-tagged USP33 and CP110, and the $SCF^{cyclin\ F}$ complex, from insect cells infected with baculoviruses[10] (Supplementary Fig. 5a, b). CP110 was ubiquitinated *in vitro* by the $SCF^{cyclin\ F}$ complex specifically via a K48 ubiquitin linkage (data not shown). Ubiquitinated CP110 was deubiquitinated by wild-type USP33 protein, but not by a catalytically inactive mutant (USP33 DM) (Fig. 3e, f). Moreover, USP33 activity was inhibited by ubiquitin aldehyde, a DUB inhibitor. Thus, USP33 specifically deubiquitinates CP110 *in vitro* and *in vivo*, explaining how USP33 stabilizes CP110.

To study the biological importance of USP33-mediated stabilization of CP110, we investigated whether USP33 depletion affects centrosome duplication. It has been shown that cyclin-F ablation leads to CP110 upregulation, which markedly induces centrosome amplification[3]. Indeed, cyclin-F knockdown induced centrosome amplification, visualized by the acquisition of more than four centrin 2 or CP110 foci and more than two γ-tubulin foci, whereas co-depletion of USP33 with cyclin F reversed this phenotype in U2OS cells (Fig. 4a and Supplementary Fig. 6a). This phenotype was also observed in another cell line (HeLa) and was similarly reversed by treatment with distinct USP33 siRNAs (Supplementary Fig. 6b). The centrosome phenotype provoked by USP33 silencing was not due to altered cell-cycle progression (Supplementary Fig. 6c). Furthermore, we investigated whether USP33 depletion could affect centrosome re-duplication in cells treated with hydroxyurea (HU) to induce prolonged S phase. Similar to CP110 ablation[2], depletion of USP33 significantly inhibited HU-induced centrosome amplification in two cell lines (Fig. 4b and Supplementary Fig. 6d). Thus, although USP33 depletion did not affect normal centriole duplication (Fig. 4a), silencing of USP33 specifically inhibited centrosome over-duplication, reminiscent of the CP110-depletion phenotype. Importantly, USP33 co-depletion rescued the mitotic defects observed in cyclin-F-depleted cells: whereas amplified centrosomes resulting from cyclin-F depletion led to multipolar and clustered bipolar (pseudo-bipolar) asymmetric spindles, co-silencing of USP33 and cyclin F reverted the frequency of mitotic aberrations to control levels (Supplementary Fig. 6e). These results suggest that destabilization of CP110 through USP33 silencing can inhibit mitotic errors wrought by centrosomal defects.

To investigate further the role of USP33 in regulating centrosome duplication, we ectopically expressed USP33 in U2OS cells synchronized by double-thymidine block and release. We found that

©2013 Macmillan Publishers Limited. All rights reserved

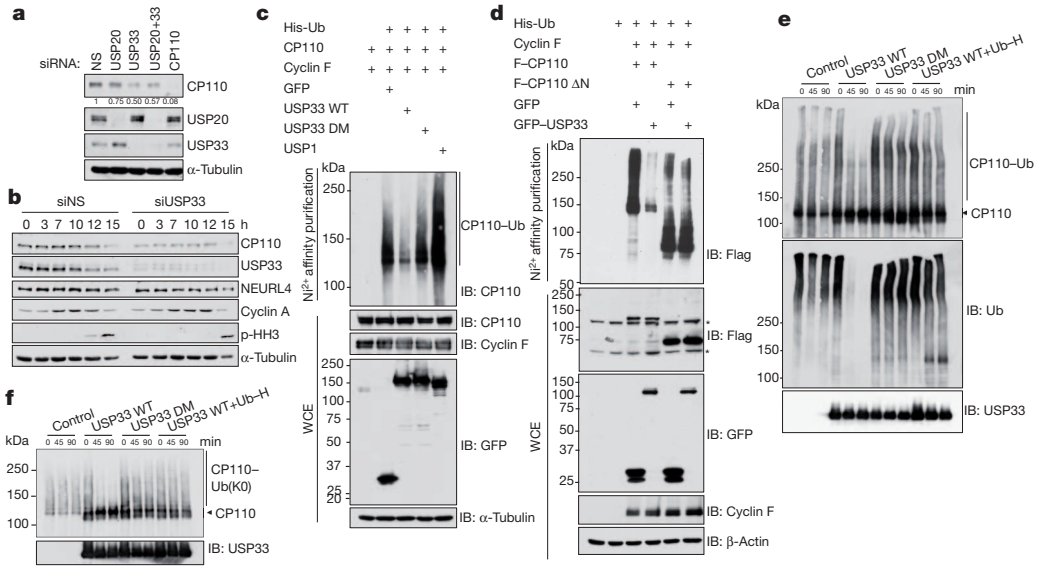


Figure 3 | Recruitment of USP33 leads to deubiquitination and stabilization of CP110. **a**, CP110 levels were monitored after silencing of USP20 and USP33 by transfection of siRNAs into HeLa cells. NS, non-specific siRNA control. Numbers indicate quantification of CP110 band intensity normalized to non-specific siRNA control. **b**, Cells were transfected with the indicated siRNAs and synchronized by double-thymidine block and release. Lysates of cells from indicated time points (hours) after release were probed with antibodies against CP110 and an interacting protein (NEURL4), as well as cell-cycle markers (cyclin A, phospho-histone H3 (Ser 10)). S10, antibody against phosphorylated Ser 10 of histone H3. **c**, Enforced expression of GFP-tagged, wild-type (WT) USP33, but not a catalytically inactive version (DM), drives deubiquitination of CP110. The indicated plasmids (top) were transfected into cells, the resulting tagged ubiquitinated proteins were enriched by nickel-NTA chelation chromatography, and western blots were probed with antibodies against CP110. An unrelated DUB (USP1) expressed as a control had no effect on CP110 ubiquitination. WCE, whole-cell extract. IB, immunoblot. **d**, Deubiquitination by USP33 requires selective binding to an amino-terminal portion of CP110. 293T cells were transfected with indicated GFP- and Flag-tagged (F) plasmids and processed as in **c**. Asterisks indicate non-specific bands. **e**, **f**, *In vitro* deubiquitination reactions were performed using wild-type ubiquitin (Ub) (**e**) or a mutant form of ubiquitin (Ub-K0) in which all lysines have been mutated (**f**), and analysed by western blotting with the indicated antibodies. **g**, The ubiquitin mutant (Ub-K0) lacks the ability to form poly-ubiquitin chains and can only be conjugated to substrate as mono-ubiquitin. *In vitro* deubiquitination of CP110 was blocked by simultaneous incubation with ubiquitin aldehyde (Ub-H) as indicated.

expression of green fluorescent protein (GFP)-tagged wild-type USP33 significantly increased the percentage of cells with amplified centrin 2, CP110 and γ-tubulin foci in comparison with the GFP control and USP33 DM mutant, and correlative light-electron microscopy confirmed the presence of bona fide ectopic centrioles and electron-dense structures, likely to be centriolar precursors or intermediates, in a subset of cells expressing GFP–USP33 (Fig. 4c and Supplementary Fig. 7a–f). Interestingly, we observed that the USP33 DM mutant also increased the percentage of cells with amplified centrioles. This may be due to residual activity of the mutant *in vivo*, additional, non-catalytic roles of USP33, or a combination of both. Alternatively, excessive amounts of USP33 DM might bind CP110, blocking it from accessing $SCF^{cyclin\ F}$. USP33 promotes the generation of excessive centriolar foci due to increased levels of CP110 because (1) overexpression of human or murine USP33 led to CP110 upregulation in multiple cell lines (Supplementary Fig. 7c, d), and (2) depletion of CP110 overrode USP33-mediated generation of centriolar foci (Supplementary Fig. 7g–i). In sum, increased levels of USP33 positively regulate centrosome duplication by augmenting CP110 levels, and loss of USP33 can prevent centrosome amplification and mitotic defects.

Centrosome amplification is frequently detected in human cancers, and it has been implicated as a contributing factor during cellular transformation[12,13]. Interestingly, a search of an integrated cancer microarray database (Oncomine) revealed that *USP33* gene expression is significantly upregulated in pancreatic cancer[14,15] (Supplementary Fig. 9a). This was of particular interest because centrosome amplification can be detected in primary pancreatic ductal carcinomas (PDAC)[16,17], and because a mouse model suggested that supernumerary centrosomes can promote chromosomal aberrations and pancreatic cancer progression[18]. Therefore, we examined USP33 and CP110 levels in human pancreatic cell lines and tissue samples. Remarkably, USP33 and CP110 were both substantially upregulated in cultured PDAC cell lines as compared to normal pancreatic epithelial cells, and USP33 and CP110 levels were strongly correlated (Supplementary Fig. 8a). Furthermore, depletion of USP33 in two PDAC cell lines (Panc1 and MiaPaCa2) led to a reduction in CP110 levels (Supplementary Figs 4d and 6d), indicating that USP33 stabilizes CP110 in PDAC cells. Notably, CP110 and USP33 levels also correlated with centrosome amplification in pancreatic cells (Supplementary Fig. 8a, b), and we found that USP33 and CP110 depletion inhibited prolonged S-phase centrosome amplification in MiaPaCa2 cells (Supplementary Fig. 6d).

Next, we examined USP33 and CP110 protein levels using human tissue arrays that included PDAC and histologically normal pancreatic samples. Two separate arrays were scored using a four-tier grading system. Notably, we found that both USP33 and CP110 were highly expressed in PDAC but not in normal pancreas, and the levels of USP33 and CP110 were significantly correlated in the same tissue samples (Supplementary Fig. 9b). In addition, we also visualized CP110, acetylated tubulin, γ-tubulin and keratin 19, a marker of pancreatic ducts and PDAC cells, using a tissue microarray containing 30 additional, independent cases of PDAC. In histologically normal ducts, a normal pattern of centriolar CP110 staining and ciliary acetylated tubulin staining was most frequently encountered (Supplementary Fig. 9c). In contrast, CP110 overexpression was observed in PDAC cells in 29 of 30 cases. Our results suggest that the abundance of centrosome regulators, such as USP33 and CP110, may be mis-regulated in human tumours that are associated with centrosomal defects, although additional studies will be needed to elucidate the temporal connections between de-regulated expression of these proteins, centrosome aberrations, and PDAC pathogenesis.

We have identified USP33 as a centrosomal DUB that stabilizes CP110 at centrioles through deubiquitination. USP33 positively

©2013 Macmillan Publishers Limited. All rights reserved

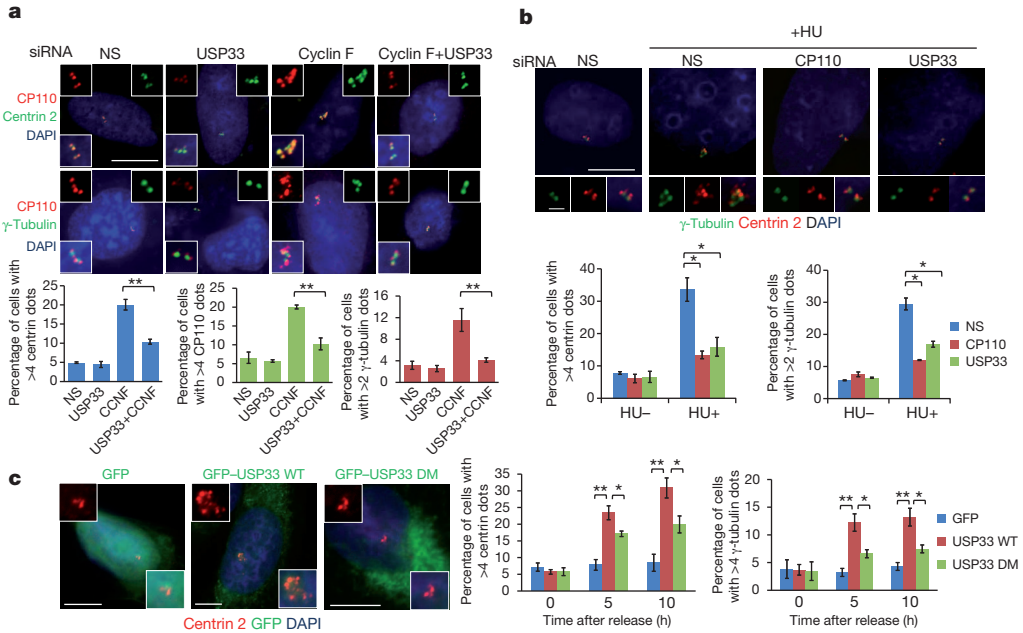


Figure 4 | Depletion of USP33 inhibits centrosome amplification and antagonizes cyclin-F function. **a**, After cyclin-F knockdown, cells express small, but measurable, amounts of residual cyclin F, the activity of which may be largely counteracted by USP33. Top, silencing of USP33 reverses the over-duplication of centrosomes observed on cyclin-F knockdown. Immunofluorescence detection of centrin (top, green), CP110 (red) and γ-tubulin (bottom, green) in U2OS cells depleted of USP33, cyclin F, or both USP33 and cyclin F. Bottom, quantification of results in top panels. NS, non-specific siRNA control. Scale bar, 10 μm. Error bars indicate standard error of the mean (s.e.m.). $^{**}P < 0.01$ ($n = 3$). **b**, Top, cells were depleted of USP33 and treated with HU. USP33 depletion suppresses HU-mediated centrosome amplification in U2OS cells. Bottom, quantification of centrin-2 and γ-tubulin foci shown in top panels. Error bars indicate s.e.m. $^{*}P < 0.05$ ($n = 3$). **c**, Cells were transfected with the indicated plasmids and synchronized with double-thymidine block and release. GFP and centrin-2 fluorescence was visualized (left), and populations of cells with the indicated number of centrin-2 and γ-tubulin dots were counted (right). Scale bar, 10 μm. Error bars indicate s.e.m. $^{*}P < 0.05$; $^{**}P < 0.01$ ($n = 3$).

regulates centrosome duplication by regulating CP110 abundance. Together with the finding that $SCF^{cyclin\ F}$ promotes the ubiquitination and degradation of CP110 (ref. 3), we have uncovered an underlying cycle in which the equilibrium between ubiquitination and deubiquitination governs the levels of a critical centrosome protein during the cell cycle, thereby ensuring centrosome homeostasis. Disruption of either the deubiquitinating or ubiquitinating enzymes could upset this equilibrium and lead to centrosome aberrations. Ubiquitinating and deubiquitinating enzymes have been the subject of intense scrutiny in cancer research, and inhibitors of the enzymes that promote or remove ubiquitin, including USP33, are targets for drug development[12,19,20]. As many human tumours are associated with centrosome amplification and genomic instability, our findings suggest the possibility of therapeutic intervention in cancers with this hallmark.

METHODS SUMMARY

Plasmids. Human and murine USP33 complementary DNAs were provided by G. Wu[6,7] and S. Shenoy[11]. USP33 mutants were generated by the QuikChange Site-directed Mutagenesis kit (Stratagene). CP110 and cyclin-F constructs have been described[3].

Biochemical methods. Extract preparation, immunoprecipitation and immunoblotting have been described[1–4].

Immunofluorescence microscopy. Indirect immunofluorescence procedures have been described[1–3]. Briefly, cells were fixed with cold methanol for 2 min or with 10% formalin solution (Sigma-Aldrich) for 8 min, followed by permeabilization with PBS and 1% Triton X-100 for 5 min and blocking in PBS, 0.1% Triton X-100 and 3% BSA for 1 h, before incubation with primary antibodies for 2 h. Cy3-conjugated donkey anti-rabbit IgG (Jackson Immunolabs) or Alexa-Fluor-488-conjugated donkey anti-mouse IgG (Invitrogen) were used as secondary antibodies. DNA was stained with DAPI. Prolong Gold (Invitrogen) was used to mount slides. Images were acquired with a Zeiss Axiovert 200 M microscope (×63 objective lens, N.A. 1.4, 1.63 Optovar), equipped with a cooled Retiga 2000R CCD (QImaging). Some images represent the maximum projection of 12 deconvoluted planes using Metamorph (Molecular Devices).

Full Methods and any associated references are available in the online version of the paper.

Received 5 April 2012; accepted 29 January 2013.

1. Chen, Z., Indjeian, V. B., McManus, M., Wang, L. & Dynlacht, B. D. CP110, a cell cycle-dependent CDK substrate, regulates centrosome duplication in human cells. *Dev. Cell* **3,** 339–350 (2002).
2. Spektor, A., Tsang, W. Y., Khoo, D. & Dynlacht, B. D. Cep97 and CP110 suppress a cilia assembly program. *Cell* **130,** 678–690 (2007).
3. D'Angiolella, V. *et al.* $SCF^{Cyclin\ F}$ controls centrosome homeostasis and mitotic fidelity through CP110 degradation. *Nature* **466,** 138–142 (2010).
4. Li, J. *et al.* Neurl4, a novel daughter centriole protein, prevents formation of ectopic microtubule organizing centres. *EMBO Rep.* **13,** 547–553 (2012).
5. Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138,** 389–403 (2009).
6. Li, Z. *et al.* Ubiquitination of a novel deubiquitinating enzyme requires direct binding to von Hippel-Lindau tumor suppressor protein. *J. Biol. Chem.* **277,** 4656–4662 (2002).
7. Li, Z. *et al.* Identification of a deubiquitinating enzyme subfamily as substrates of the von Hippel–Lindau tumor suppressor. *Biochem. Biophys. Res. Commun.* **294,** 700–709 (2002).
8. Curcio-Morelli, C. *et al.* Deubiquitination of type 2 iodothyronine deiodinase by von Hippel-Lindau protein-interacting deubiquitinating enzymes regulates thyroid hormone activation. *J. Clin. Invest.* **112,** 189–196 (2003).
9. Thorne, C., Eccles, R. L., Coulson, J. M., Urbe, S. & Clague, M. J. Isoform-specific localization of the deubiquitinase USP33 to the Golgi apparatus. *Traffic* **12,** 1563–1574 (2011).
10. D'Angiolella, V. *et al.* Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair. *Cell* **149,** 1023–1034 (2012).
11. Berthouze, M., Venkataramanan, V., Li, Y. & Shenoy, S. K. The deubiquitinases USP33 and USP20 coordinate β_2 adrenergic receptor recycling and resensitization. *EMBO J.* **28,** 1684–1696 (2009).
12. Chan, J. Y. A clinical overview of centrosome amplification in human cancers. *Int. J. Biol. Sci.* **7,** 1122–1144 (2011).
13. Ganem, N. J., Godinho, S. A. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460,** 278–282 (2009).
14. Buchholz, M. *et al.* Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. *Oncogene* **24,** 6626–6636 (2005).
15. Badea, L., Herlea, V., Dima, S. O., Dumitrascu, T. & Popescu, I. Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal

©2013 Macmillan Publishers Limited. All rights reserved

adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepatogastroenterology* **55,** 2016–2027 (2008).
16. Sato, N. *et al.* Centrosome abnormalities in pancreatic ductal carcinoma. *Clin. Cancer Res.* **5,** 963–970 (1999).
17. Sato, N. *et al.* Correlation between centrosome abnormalities and chromosomal instability in human pancreatic cancer cells. *Cancer Genet. Cytogenet.* **126,** 13–19 (2001).
18. Shono, M. *et al.* Stepwise progression of centrosome defects associated with local tumor growth and metastatic process of human pancreatic carcinoma cells transplanted orthotopically into nude mice. *Lab. Invest.* **81,** 945–952 (2001).
19. Cohen, P. & Tcherpakov, M. Will the ubiquitin system furnish as many drug targets as protein kinases? *Cell* **143,** 686–693 (2010).
20. Sippl, W., Collura, V. & Colland, F. Ubiquitin-specific proteases as cancer drug targets. *Future Oncol.* **7,** 619–632 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Hajdu, G. Miller, L. Chiriboga, H. Court, C. Zambirinis, E. Bekes, L. Taylor, S. Krauter, M. Philips and D. Bar-Sagi for help with staining and analysis of pancreatic cancer samples. We thank F. Liang, Y. Deng, C. Petzold and K. Dancel for help with correlative electron microscopy. We thank J. Jastrab and S. Vijayakumar for initial experiments pertaining to the USP20–CP110 interaction. We thank J. Pollack for providing pancreatic cell lines. We thank G. Wu, S. Shenoy, R. Baer and T. Huang for providing plasmids. We thank members of the Dynlacht laboratory and T. Huang for advice. B.D.D. was supported by National Institutes of Health grant 5R01HD069647-02 and March of Dimes grant FY11-432. M.P. is an Investigator with the Howard Hughes Medical Institute.

Author Contributions J.L., V.D'A., S.K., T.K., E.I.C., M.P. and B.D.D. performed all biochemical experiments and analyses of deubiquitination and centrosome phenotypes. E.S.S., W.F. and J.L. performed pancreatic tissue studies. J. L. and B.D.D. wrote the manuscript. B.D.D. coordinated the study and oversaw all experiments. All authors discussed the results and commented on and assisted in revising the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.D.D. (Brian.Dynlacht@nyumc.org).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Cell culture and cell-cycle synchronization. HeLa, U2OS, RPE1-hTERT, HEK-293T, COS-1 and COS-7 were obtained from ATCC and maintained according to their instruction. Pancreatic cell lines were provided by J. Pollack. Early passages of cells were used, and they were not constitutively cultured for more than 2 months. For cell-cycle synchronization, HeLa or U2OS cells were cultured in the presence of 2 mM thymidine (Sigma) for 16 h, washed twice with medium, and released into fresh medium without thymidine for 8 h. After another 16 h block in thymidine, cells were washed twice, released to fresh medium, and collected at indicated time points (5 h, S phase; 10 h, G2 phase). Nocodazole (100 ng ml^{-1}) was added 5 h after release from the thymidine block to block cells in pro-metaphase.

Transient transfections. siRNAs were transfected into HeLa, U2OS or RPE-1 cells using siIMPORTER (Millipore), HiPerFect (Qiagen) or Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Plasmids were transfected into HeLa and U2OS cells using FuGene6 (Roche) or Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. HEK-293T, COS-1 and COS-7 cells were transfected using the calcium phosphate method, as described[1–3].

Antibodies. Rabbit polyclonal antibodies against CP110 (ref. 1), NEURL4 (ref. 4) and CEP97 (ref. 2) have been described. Immunofluorescence staining of CP110 was performed with CP110 antibody from Proteintech (12780-1-AP). Antibodies against USP33 (A300-925A; Bethyl Laboratories; 1:100 dilution for immunofluorescence), USP20 (A301-189A; Bethyl Laboratories), centrin 2 (20H5; millipore), cyclin F (C-20; Santa Cruz Biotechnology), Flag (M2; Sigma), α-tubulin (T5168; Sigma), γ-tubulin (mouse, T5326; Sigma), γ-tubulin (rabbit, T5192; Sigma), β-actin (AC-15; abcam), ubiquitin (P4D1, Enzo life sciences and FK2, Biomol International), GFP (rabbit, G1544; Sigma), GFP (mouse, Roche), cyclin A (C-19; Santa Cruz Biotechnology), CEP290 (A301-659A; Bethyl Laboratories), phosphorylated histone H3 (phospho-Ser 10; Millipore), phosphorylated CDC2 (phospho-Tyr15; Santa Cruz Biotechnology) were also used.

siRNA sequence. The 21-nucleotide siRNA sequence for the non-specific control was AATTCTCCGAACGTGTCACGT[2]. The sequences for USP33 oligonucleotides 1 and 2 were GAUCAUGUGGCGAAGCAUAdTdT and GGCUUGGA UCUUCAGCCAUdTdT[11], respectively. The siRNA for CP110 was AAGCAG CATGAGTATGCCAGT[2]. The siRNA sequence for USP20 was GGACAAUGA UGCUCACCUAUU[11]. The siRNA sequence for NEURL4 was CCAUCAUG CAAGACGGUAAUU[4].

***In vivo* deubiquitination assay.** The methods to detect ubiquitinated protein were modified from ref. 21. HEK293 cells were transiently transfected with pcB6-His ubiquitin (gift from R. Baer), HA- or Flag- tagged CP110, cherry-cyclin F, GFP, GFP–USP33 wild type, GFP–USP33 catalytic mutant, or GFP–USP1 (a gift from T. Huang), as indicated. After 36 h, cells were treated with 10 μm MG132 (Peptides International) for 4 h, cells were lysed and briefly sonicated in Buffer A (100 mM NaH_2PO_4, 10 mM Tris-Cl, 6 M guanidine-HCl, 10 mM imidazole pH 8.0). The cell lysates were incubated with nickel-NTA resin (Qiagen) to precipitate His-tagged ubiquitinated proteins for 3 h at room temperature (24 °C). Ni-NTA beads were subjected to two washes with buffer A, two washes with buffer A/TI (1 volume buffer A and 1 volume TI buffer) (TI buffer, 25 mM Tris-Cl pH 6.8 and 20 mM imidazole), and then one wash with TI buffer. Finally, Ni-NTA beads were eluted in SDS loading buffer containing 200 mM imidazole, separated by SDS–PAGE, and detected by immunoblotting.

***In vitro* deubiquitination assay.** Flag-tagged USP33 and CP110 and the SCF$^{\text{cyclin F}}$ complex (comprising GST–SKP1, His-tagged CUL1, ROC1 and His-tagged cyclin F) were individually affinity purified using anti-Flag (CP110 and USP33) or glutathione-agarose (SCF$^{\text{cyclin F}}$) chromatography from extracts of baculovirus-infected insect cells[10]. CP110 was ubiquitinated *in vitro* by purified SCF$^{\text{cyclinF}}$ complex in ubiquitination reaction buffer (50 mM Tris-HCl pH 7.6, 5 mM $MgCl_2$, 1 μM okadaic acid and 0.6 mM dithiothreitol (DTT)) containing 2 mM ATP, 0.1 μM E1 (Boston Biochem),10 ng μl^{-1} Ubch3, 10 μg ml^{-1} Ubch5c and 2.5 mg ml^{-1} ubiquitin wild type or ubiquitin-k0 (Boston Biochem). Ubiquitinated CP110 was then diluted tenfold with deubiquitination reaction buffer (20 mM HEPES pH 7.6, 50 mM NaCl, 1 mM EDTA, 0.2 μg μl^{-1} BSA, 1 mM DTT, 100 μM MG132, 1 μM AEBSF, 2 μg ml^{-1} aprotinin). Diluted ubiquitinated CP110 was incubated with 10 ng μl^{-1} USP33 protein at 30 °C for the indicated time (0–1.5 h). For the reactions with ubiquitin aldehyde, 8 μM ubiquitin aldehyde (Boston Biochem) was added. Samples were separated by SDS–PAGE, and detected by immunoblotting.

Correlative light-electron microscopy. U2OS cells were plated at low confluence on gridded glass bottom dishes (MatTek) and transfected with GFP–USP33/RFP–centrin 2 or GFP along/RFP–centrin 2 using FuGene 6. After 48 h, cells were examined by live-cell imaging to monitor green and red fluorescence and to obtain phase contrast images. Cells with both GFP and RFP expression were documented, and their relative positions on the grids were recorded. Cells were then fixed with 0.1 M sodium cacodylate buffer (pH 7.4) supplemented with 2.5% glutaraldehyde, and 1 mM $CaCl_2$, and then processed in a standard manner for transmission electron microscopy, as previously described[4].

Tissue microarray. Pancreatic tissue arrays were purchased from US Biomax (PA1001 and PA207). Immunohistochemical staining was performed as described[22] and scored by two independent pathologists using a four-tier grading system (0 = negative, 1 = weak, 2 = moderate and 3 = strong staining intensity). Immunofluorescent staining was performed as described[23].

Statistical analyses. All data represent the average from at least three independent experiments with at least 100 cells counted per experiment unless otherwise indicated. Significance was calculated by two-tailed Student's *t*-test or analysis of variance (Supplementary Fig. 6e) using GraphPad software. Differences were considered significant when $P < 0.05$.

21. Tansey, W. P. Detection of ubiquitylated proteins in mammalian cells. *Cold Spring Harb. Protoc.* http://dx.doi.org/10.1101/pdb.prot4616 (2006).
22. Sabbaghian, M. S. *et al.* Levels of elevated circulating endothelial cell decline after tumor resection in patients with pancreatic ductal adenocarcinoma. *Anticancer Res.* **30,** 2911–2917 (2010).
23. Seeley, E. S., Carriere, C., Goetze, T., Longnecker, D. S. & Korc, M. Pancreatic cancer and precursor pancreatic intraepithelial neoplasia lesions are devoid of primary cilia. *Cancer Res.* **69,** 422–430 (2009).

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature11899

Crystal structures of the calcium pump and sarcolipin in the Mg^{2+}–bound E1 state

Chikashi Toyoshima[1], Shiho Iwasawa[1], Haruo Ogawa[1], Ayami Hirata[1], Junko Tsueda[1] & Giuseppe Inesi[2]

P-type ATPases are ATP-powered ion pumps that establish ion concentration gradients across biological membranes, and are distinct from other ATPases in that the reaction cycle includes an autophosphorylation step. The best studied is Ca^{2+}-ATPase from muscle sarcoplasmic reticulum (SERCA1a), a Ca^{2+} pump that relaxes muscle cells after contraction, and crystal structures have been determined for most of the reaction intermediates[1,2]. An important outstanding structure is that of the E1 intermediate, which has empty high-affinity Ca^{2+}-binding sites ready to accept new cytosolic Ca^{2+}. In the absence of Ca^{2+} and at pH 7 or higher, the ATPase is predominantly in E1, not in E2 (low affinity for Ca^{2+})[3], and if millimolar Mg^{2+} is present, one Mg^{2+} is expected to occupy one of the Ca^{2+}-binding sites with a millimolar dissociation constant[4,5]. This Mg^{2+} accelerates the reaction cycle[4], not permitting phosphorylation without Ca^{2+} binding. Here we describe the crystal structure of native SERCA1a (from rabbit) in this E1·Mg^{2+} state at 3.0 Å resolution in addition to crystal structures of SERCA1a in E2 free from exogenous inhibitors, and address the structural basis of the activation signal for phosphoryl transfer. Unexpectedly, sarcolipin[6], a small regulatory membrane protein of Ca^{2+}-ATPase[7], is bound, stabilizing the E1·Mg^{2+} state. Sarcolipin is a close homologue of phospholamban, which is a critical mediator of β-adrenergic signal in Ca^{2+} regulation in heart (for reviews, see, for example, refs 8–10), and seems to play an important role in muscle-based thermogenesis[11]. We also determined the crystal structure of recombinant SERCA1a devoid of sarcolipin, and describe the structural basis of inhibition by sarcolipin/phospholamban. Thus, the crystal structures reported here fill a gap in the structural elucidation of the reaction cycle and provide a solid basis for understanding the physiological regulation of the calcium pump.

New E2 (Supplementary Fig. 1) crystals of SERCA1a free from exogenous inhibitors were grown in either 40 mM $MgSO_4$ (E2·SO_4^{2-} crystals) or $MgCl_2$ (E2 crystals) at pH 6.5 and diffracted to 2.45 or 2.5 Å resolution. The structures of the ATPase as well as the crystal packing are virtually identical to those with thapsigargin[12,13] (root mean squared deviation (r.m.s.d.) is 0.75 Å) (Fig. 1 and Supplementary Fig. 2). This indicates that thapsigargin simply fits into a pre-formed cavity in E2 and locks the structure[14]. One sulphate ion was located near the phosphorylation site (Asp 351) in E2·SO_4^{2-}, at a position different from that of the inorganic-phosphate analogue (MgF_4^{2-}) in E2·MgF_4^{2-}(TG) (ref. 15, where TG indicates thapsigargin, Supplementary Fig. 3) but identical to that of SO_4^{2-} in CopB, a copper ATPase[16], possibly representing a phosphate ion before release into the bulk medium.

At pH 7.5, crystals of SERCA1a grew only in the presence of 40 mM $MgSO_4$ (E1·Mg^{2+} crystals); those from sarcoplasmic reticulum diffracted to 3.0 Å and those of recombinant SERCA1a to 3.2 Å resolution. Addition of 0.3 mM trinitrophenyl adenosine monophosphate (TNPAMP) improved the resolution from 3.5 to 3.0 Å (Supplementary Table 1). TNPAMP crosslinks the N domain and P domain[17], and slightly changed the orientation of the N domain but hardly affected any other part of the structure (r.m.s.d. = 0.44 Å excluding the N domain; Supplementary Fig. 4). Therefore, the TNPAMP-bound form is described here for E1·Mg^{2+}.

SERCA1a is composed of a large cytoplasmic headpiece consisting of the A (actuator), N (nucleotide binding) and P (phosphorylation) domains, 10 transmembrane helices (M1–M10) and small lumenal loops. The most conspicuous structural change caused by Ca^{2+}-binding as revealed by E1·$2Ca^{2+}$ and E2 crystal structures is the appearance of the headpiece (Fig. 1)[1]. It is wide open in E1·$2Ca^{2+}$, and closed and compact in E2 so that the A domain stays over the phosphorylation site to prevent phosphoryl transfer from ATP. In the transition into E1·$2Ca^{2+}$, the A domain rotates by ~110° around an axis approximately perpendicular to the membrane to expose the phosphorylation site. The rearrangement in the transmembrane domain is also drastic: M4 moves towards the cytoplasm, together with M3, to form Ca^{2+}-binding site II; M5 straightens to open the headpiece by separating the N and A domains; the unwound part of M6 rotates to form site I. M1 is sharply bent in E2 but now straightens; M2 is disrupted in E2 but contiguous in E1·$2Ca^{2+}$ (see ref. 1 for a full account).

The crystal structure of Ca^{2+}-ATPase in E1·Mg^{2+} is intermediate between E2 and E1·$2Ca^{2+}$ but much closer to E1·$2Ca^{2+}$ (Fig. 1), explaining the accelerating effects on Ca^{2+}-binding of Mg^{2+}. For instance, the headpiece is indeed open, with the A domain completely detached from the N domain and azimuthally rotated by ~110° from E2 (rotation 1), very similar to that in E1·$2Ca^{2+}$. The only conspicuous difference is the M1 helix, which is sharply bent as in E2 and occupies an even lower position with respect to the membrane. As a result, a large open 'mouth' forms that leads to the transmembrane Ca^{2+}-binding sites (Supplementary Fig. 5).

Such changes in structure arise from H^+ release and Mg^{2+} binding, the events that take place in the transmembrane Ca^{2+}-binding sites. In the $|F_{obs}| - |F_{calc}|$ electron density map, there is a strong peak at ~5.5σ with an elongated shape in between Ca^{2+}-binding sites I and II, but much closer to (2.7 Å from) site II (Fig. 2b). Although it is impossible at this resolution (3.0 Å) to unambiguously identify ion species or precise coordination geometry, the most likely interpretation is to place one Mg^{2+} at the centre with one or two associated water molecules (Supplementary Fig. 6 for an electron density map). Preliminary all-atom molecular dynamics simulation identifies two water molecules bound to the Mg^{2+} at these positions. Here four side-chain oxygen atoms, Ala 305 main-chain carbonyl, and the putative water molecule(s) contribute to make a 6 (or 7) coordination expected for Mg^{2+} binding[18]. Nevertheless, distances between the Mg^{2+} and the oxygen atoms are longer than ideal (2.0–2.2 Å), consistent with the millimolar K_d of Mg^{2+}-binding in this state[4,5]. The Ca^{2+}-binding site I is formed but vacant, because the oxygen atoms of the carboxyl groups (those of Glu 771 and Asp 800) that coordinate Mg^{2+} are not used for Ca^{2+}-binding at site I in E1·$2Ca^{2+}$ (ref. 1). Glu 908, another site I residue (Fig. 2c), is not engaged in Mg^{2+}-binding (Fig. 2b). Furthermore, there is plenty of space above the bound Mg^{2+} for Ca^{2+} to reach the binding site (Supplementary Fig. 5b). As the geometry is incompatible with simultaneous binding of Ca^{2+} and Mg^{2+}, Ca^{2+} binding to site I will displace the Mg^{2+} and result in proper

[1]Institute of Molecular and Cellular Biosciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan. [2]California Pacific Medical Center Research Institute, San Francisco, California 94107, USA.

©2013 Macmillan Publishers Limited. All rights reserved

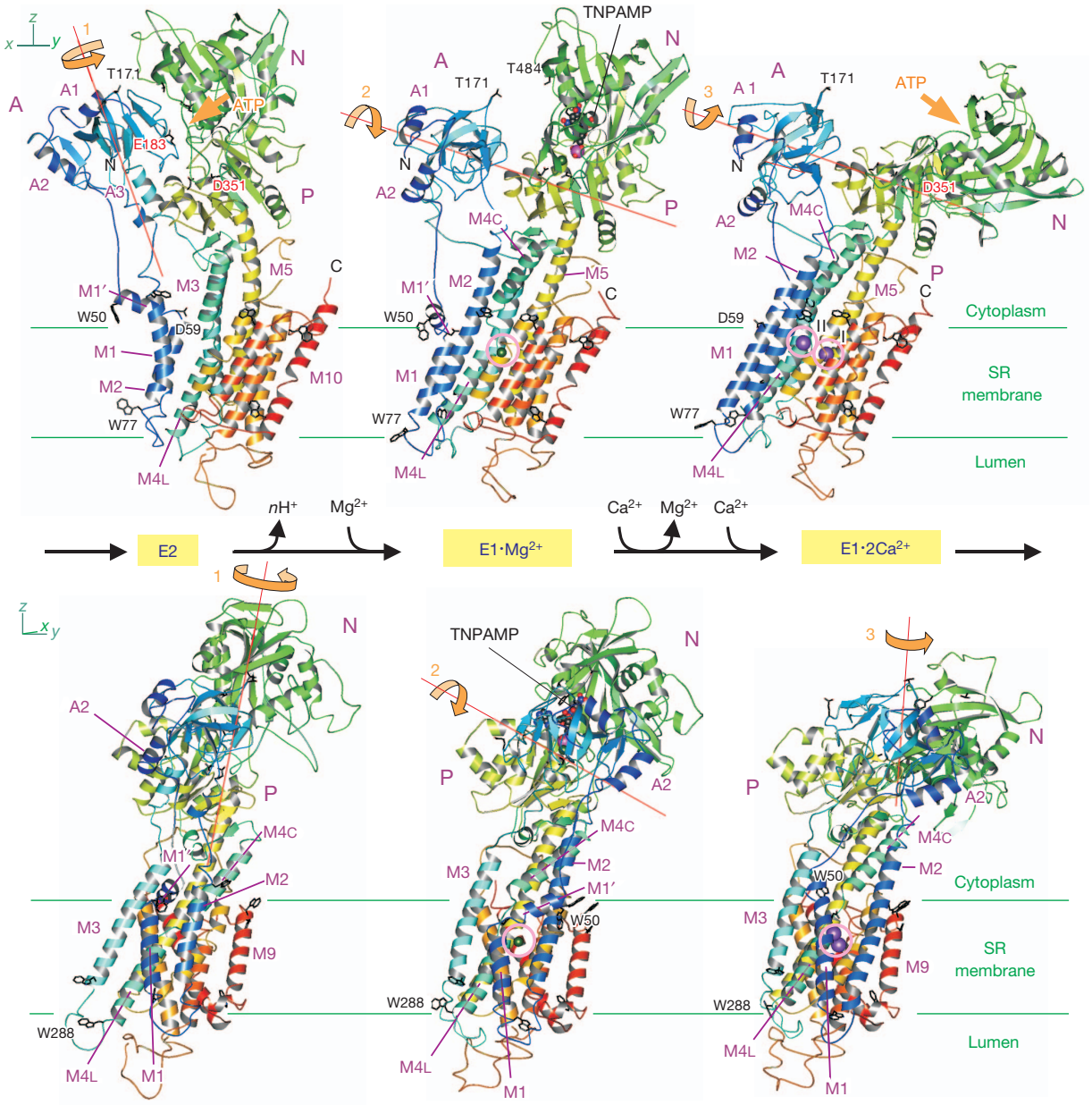


Figure 1 | Crystal structure of SERCA1a in E1·Mg^{2+} in comparison with those in E2 and E1·$2Ca^{2+}$. Structures are viewed in two orthogonal directions: upper panels, front views; lower panels, viewed from the right-hand side. The atomic model used for E1·$2Ca^{2+}$ is 1SU4. Models are aligned with the M7–M10 helices. Colours change gradually from the amino terminus (blue) to the carboxy terminus (red). Green horizontal lines indicate the approximate boundary of the hydrophobic core of the lipid bilayer. The axis of rotation of the A domain with respect to the P domain (axes 1–3, orange) is shown for the transition going to the next step as determined with Dyndom[30]. SR, sarcoplasmic reticulum.

formation of Ca^{2+}-binding site II. In E1·Mg^{2+}, site II is completely disrupted, because M4L, the lumenal half of the M4 helix, which provides three main-chain oxygen atoms for coordination of site II Ca^{2+} in E1·$2Ca^{2+}$, is shifted towards M3 (~4 Å) for coordinating Mg^{2+}, away from M6 (Fig. 2c). Thus, the high-affinity Ca^{2+}-binding sites are only half formed in E1·Mg^{2+}.

The arrangement of the transmembrane helices is unique among all the crystal structures of SERCA1a in different states (Fig. 2a), because the main-chain carbonyl of Ala 305, located at the cytoplasmic end of M4L, coordinates Mg^{2+} in E1·Mg^{2+} (Fig. 2b) but site II Ca^{2+} in E1·$2Ca^{2+}$ (Fig. 2c). As site II is located closer to the cytoplasmic surface of the membrane than site I (that is, higher in Fig. 1), Ala 305, and therefore the entire M4L, is located higher up in E1·$2Ca^{2+}$ than E1·Mg^{2+}. Reflecting the different disposition of M4L, the ^{308}PEGL loop containing Glu 309, a gating residue[1], takes a distinct path (6.2 Å distance at Gly 310 from E1·$2Ca^{2+}$) and so too the cytoplasmic half of the M4 helix (M4C) (Fig. 1).

These differences due to M4L in turn cause small but presumably critical differences in the positions of the A domain and the M1/M2 helices. In fact, with respect to the P domain, the A domain in E1·Mg^{2+} is ~30° rotated around an axis approximately horizontal compared to that in E1·$2Ca^{2+}$ (rotation 2 in Fig. 1). The A domain is also more inclined towards the membrane, reflecting the inclination of the P domain. As a result, the M1/M2 V-shaped structure is pushed downwards, making the vestibule to the Ca^{2+}-binding sites much larger than that in E2 (Supplementary Fig. 5). The functional implications of these differences will become apparent later.

In one of the protomers that constitute the asymmetric unit of the E1·Mg^{2+} crystal, there was an extra electron density in the groove surrounded by transmembrane helices M2, M6 and M9 (Supplementary

©2013 Macmillan Publishers Limited. All rights reserved

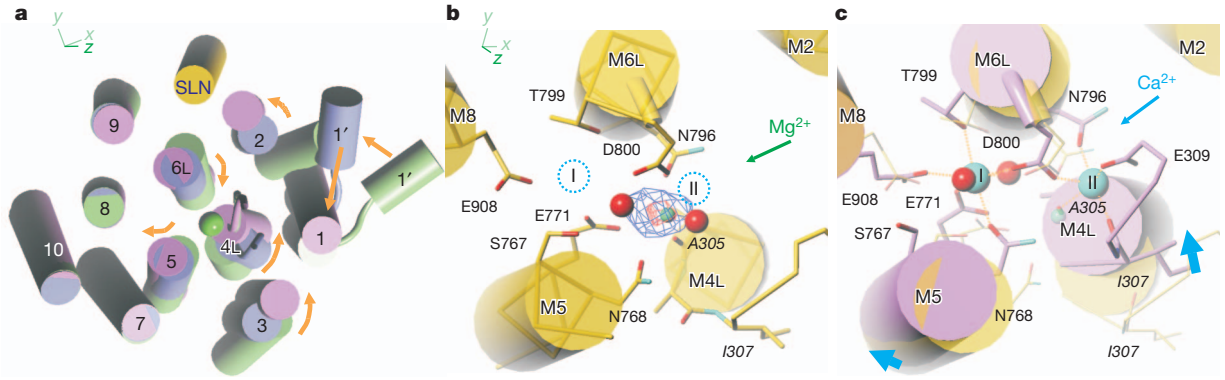


Figure 2 | Transmembrane regions in E1·Mg^{2+} and E1·$2Ca^{2+}$ crystal structures. **a**, Arrangement of transmembrane helices (cylinders). **b**, Details around the Mg^{2+} binding site in E1·Mg^{2+}. **c**, Superimposition of atomic models for E1·$2Ca^{2+}$ (pink sticks) and E1·Mg^{2+} (yellow sticks). Viewed from the cytoplasmic side. Colours used for cylinders are: lime, E2; blue grey (**a**) or yellow (**b**, **c**), E1·Mg^{2+}; pink, E1·$2Ca^{2+}$. Blue and red nets in **b** represent $|F_{obs}| - |F_{calc}|$ electron density map contoured at 3σ and 4.5σ, respectively, before introducing Mg^{2+} into the atomic model. Dotted circles in **b** indicate the positions of Ca^{2+} ions in E1·$2Ca^{2+}$. Small green spheres represent Mg^{2+}, cyan ones Ca^{2+}, red ones water. Arrows in **a** and **c** show the movements in the transition from E2 → E1·$2Ca^{2+}$ or E1·Mg^{2+} → E1·$2Ca^{2+}$. Residue names in italics in **b** and **c** indicate that the main chain carbonyl contributes to the co-ordination. Dotted lines in **c** show coordination of Ca^{2+}.

Fig. 7a). From its length and position, we thought that this extra density would represent sarcolipin, a polypeptide of 31 residues[6] present in both fast-twitch skeletal and atrial cardiac muscles (Fig. 3a; see Supplementary Fig. 8 for an electron density map). Sarcolipin is a close homologue of phospholamban (50% identity for the transmembrane part), which is the major target of the β-adrenergic signal in regulation of cardiac contractility[8,10]. Phospholamban kinetically decreases the apparent Ca^{2+} affinity of SERCA[19,20], and thereby "reserve" contractility for β-adrenergic control, which regulates phosphorylation of phospholamban that relieves its inhibitory effects[8–10]. A similar mechanism is now postulated for sarcolipin with Thr 5 as the phosphorylation site[21,22].

This groove surrounded by transmembrane helices M2, M6 and M9 is the proposed location for phospholamban/sarcolipin[23,24], as Val 49 of phospholamban is crosslinked to Val 89 on M2 of SERCA1a in

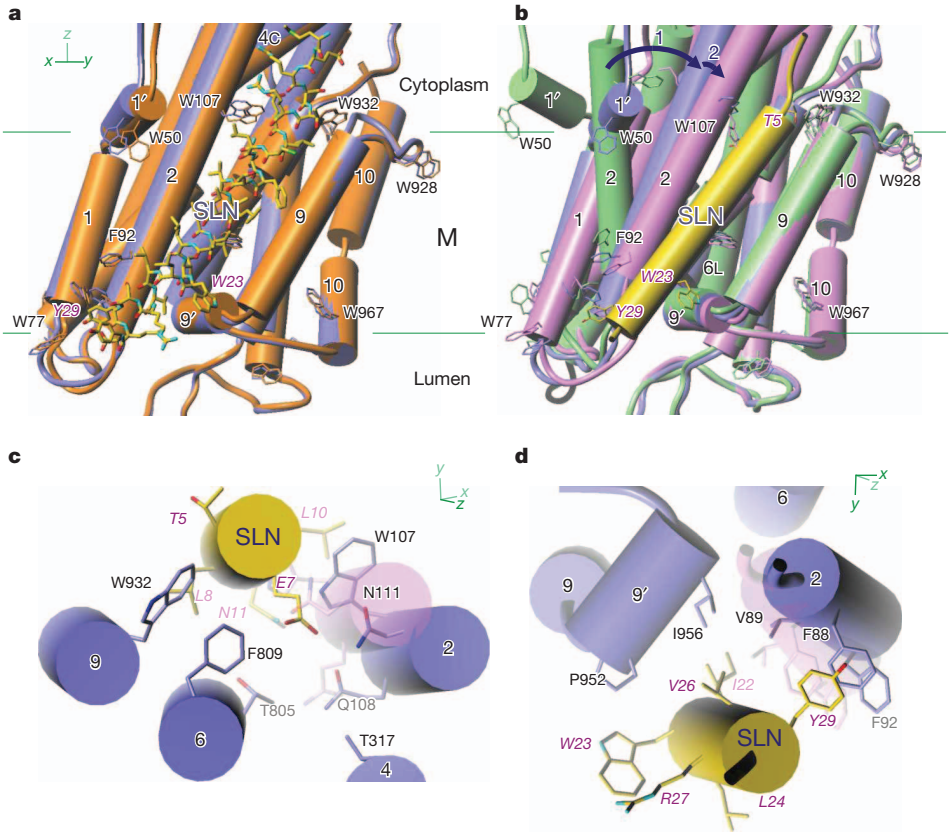


Figure 3 | Sarcolipin and its binding to SERCA1a. **a**, Transmembrane domain of SERCA1a in E1·Mg^{2+} with (native enzyme) and without (recombinant) bound sarcolipin (SLN). **b**, Rearrangement of transmembrane helices that form sarcolipin binding cavity. **c**, **d**, Details of interaction between sarcolipin and SERCA1a near the cytoplasmic (**c**) and lumenal (**d**) surface of the membrane. Colours of the cylinders representing transmembrane helices are: blue grey, E1·Mg^{2+} (native enzyme with bound sarcolipin); orange, E1·Mg^{2+} (recombinant enzyme devoid of sarcolipin); lime, E2; pink, E1·$2Ca^{2+}$. Blue arrows in **b** show the movements of M2 in E2 → E1·Mg^{2+} (arrow 1) and E1·Mg^{2+} →E1·$2Ca^{2+}$ (arrow 2). Also see Supplementary Video 1. 'M' refers to the sarcoplasmic reticulum membrane (as in Fig. 1). Residue names written in italics belong to SLN.

©2013 Macmillan Publishers Limited. All rights reserved

Cys-substituted mutants[24,25]. We carried out a series of similar crosslinking experiments and confirmed that the binding sites of sarcolipin and phospholamban indeed overlap[11,23] and that E1·Mg^{2+} is the preferred state for their binding to SERCA (Supplementary Figs 9 and 10). It was previously postulated that sarcolipin/phospholamban binds to SERCA in E2[23,24] and E2·ATP[25,26]; binding in E1·2Ca^{2+} and other states is controversial[11,25,27]. Nevertheless, there was no indication of sarcolipin in the electron density maps of E2 crystals free from exogenous inhibitors (Supplementary Fig. 7c).

To confirm further the assignment and to determine the structure of the ATPase free from sarcolipin, we crystallized recombinant ATPase expressed in COS cells. This is because our purified ATPase preparation contained sarcolipin at ~40% of the level in the original sarcoplasmic reticulum membrane (Supplementary Fig. 11), and because weak electron density was also observed with the other protomer in the asymmetric unit, which takes virtually the same conformation (r.m.s.d. = 0.43 Å). The electron density map of the recombinant ATPase under the same conditions is certainly devoid of density attributable to sarcolipin (Supplementary Fig. 7b). The overall structure is similar (r.m.s.d. = 0.95 Å) but significantly different (Fig. 3a and Supplementary Fig. 12). The A-domain rotation is intermediate between E1·2Ca^{2+} and E1·Mg^{2+} (native) (Supplementary Fig. 13a). This is presumably due to a small (~1.0 Å) movement of M2, required for opening the cavity for sarcolipin binding (Fig. 3a), which is transmitted to cytoplasmic domains through M1 and M4 (Supplementary Fig. 13), although we cannot exclude the possibility of a crystal packing effect.

The structure of sarcolipin and its interaction with Ca^{2+}-ATPase agree well with previous NMR[28] and crosslinking[11] studies. Sarcolipin, consisting of a single continuous transmembrane helix with short extensions at both ends, is not deeply embedded in the cleft surrounded by the M2, M6 and M9 helices (Fig. 3 and Supplementary Fig. 14). This cleft is particularly narrow near the cytoplasmic surface around Thr 5 of sarcolipin (~9 Å), as two Trp on M2 and M9 approach. This part becomes too narrow (<6 Å) in E1·2Ca^{2+} but too wide (>20 Å) in E2 (Fig. 3b and Supplementary Fig. 14) for sarcolipin to bind firmly, consistent with crosslinking results with Trp107Cys mutant (Supplementary Fig. 9). Thus, it is clear that sarcolipin stabilizes the E1·Mg^{2+} state and interferes with the transition into E1·2Ca^{2+} by suppressing the movement of M2 (Fig. 3b and Supplementary Videos 1 and 2). This will make the energy level of E1·Mg^{2+} lower and decrease the difference in free energy between the Ca^{2+} bound (that is, E1·2Ca^{2+}) and unbound (that is, E1·Mg^{2+}) states, and thereby decrease the apparent affinity of SERCA for Ca^{2+}.

Finally, we address why filling of two Ca^{2+} is absolutely necessary for phosphoryl transfer, which apparently requires only bending of the P domain for coordinating Mg^{2+} and ATP delivery to the phosphorylation site (Asp 351)[1]. One possibility is that the A domain blocks ATP delivery. This is true in E2, as the ^{181}TGES loop in the A domain occupies the space where the adenine ring is positioned during phosphoryl transfer (Fig. 4a)[15,29]. In E1·Mg^{2+}, the A domain is largely rotated to expose Asp 351 (rotation 1; Fig. 4a and Supplementary Video 3) but takes a position that still hinders proper approach of the N domain (Fig. 4b and Supplementary Video 4). In the transition to E1·2Ca^{2+}, the A domain undergoes a further rotation of ~30° around a different axis (rotation 2; Figs 1 and 4). Another possibility is that the bending of the P domain is forbidden because the bending pulls M1 and M2 towards cytoplasm, which is apparently allowed only from E1·2Ca^{2+}. The M1 and M2 helices form a V-shaped structure, but require differential movements of M1 (~12 Å) and M2 (~5 Å) to gain the conformation found in E1·ATP. It presumably requires stepwise movements that include straightening of the M1 helix as realized in E1·2Ca^{2+} (Fig. 4c). This straightening is an energetically costly event as the hydrophobic surface of M1′, the bent part of the M1 helix (Fig. 1), including Trp 50 becomes exposed to solvent. It is geometrically impossible in E1·Mg^{2+} because Trp 50 will collide with the ^{308}PEGL motif and because the position of the A domain appears

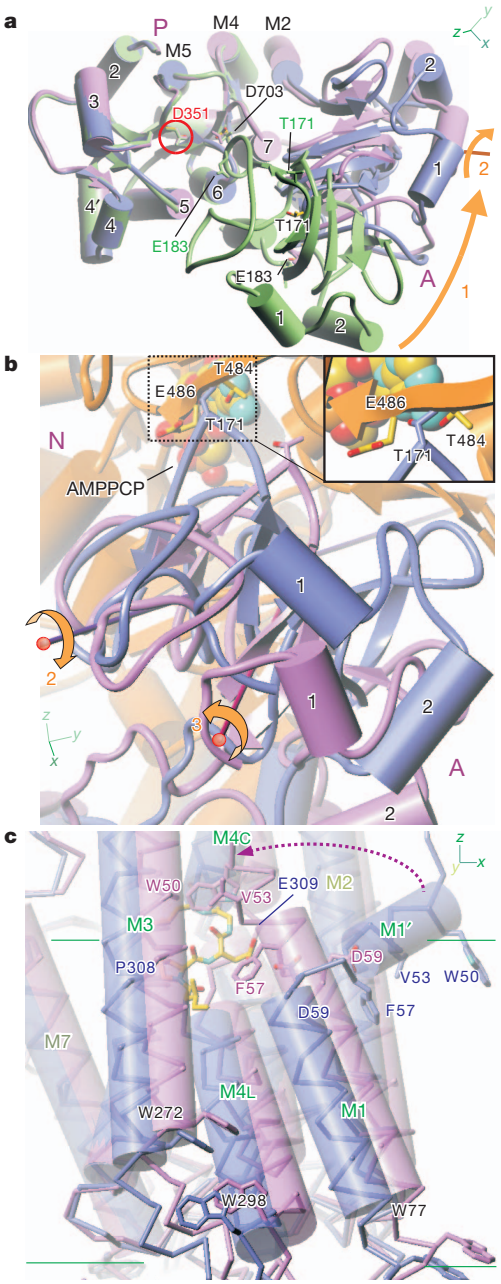


Figure 4 | Structural features of E1·Mg^{2+} that potentially prevent phosphoryl transfer before Ca^{2+}-binding. **a**, Superimposition of the A and P domains, aligned with the P domain. **b**, Superimposition of the cytoplasmic domains, aligned with the P7 helix that forms an interface with the A domain. **c**, Superimposition of the atomic models for E1·Mg^{2+} and E1·2Ca^{2+} around Trp 50, aligned with the transmembrane helices M7–M10. The A domain of E1·AMPPCP and the N and P domains of E1·Mg^{2+} and E1·2Ca^{2+} are removed in **b**. Inset in **b** shows an enlarged view of the boxed area. Colours used are: lime, E2; blue grey, E1·Mg^{2+}; pink, E1·2Ca^{2+}; orange, E1·AMPPCP (Protein Data Bank (PDB) ID: 3AR2). Orange arrows in **a** and **b** indicate the A-domain rotation with respect to the P domain in the E2 → E1·Mg^{2+} (arrow 1), E1·Mg^{2+} → E1·2Ca^{2+} (arrow 2) and E1·2Ca^{2+} → E1·ATP (arrow 3). The dotted arrow in **c** shows the movement of Trp 50 in E1·Mg^{2+} → E1·2Ca^{2+}. See also Supplementary Videos 3 and 4.

incompatible. Taken together, the A domain seems to work as a switch here, as is proposed for the lumenal gating[1]. In this regard, it is noteworthy that sarcolipin forces the A domain to undergo a larger rotation in moving into E1·2Ca^{2+} (Fig. 4a) in addition to conflicting with the movement of M2 (Fig. 3b), thus explaining the inhibitory kinetic effects of sarcolipin/phospholamban[19,20].

©2013 Macmillan Publishers Limited. All rights reserved

The E1·Mg^{2+} crystal structure reported here fills a gap in the structural elucidation of the entire reaction cycle of the calcium pump. It is surprising to see how a pump utilizes different ions to maximize efficiency and how high-affinity cation-binding sites are created stepwise in the reaction cycle. The crystal structure shows (for the first time, to our knowledge) sarcolipin in complex with SERCA and will stimulate further studies on sarcolipin and phospholamban, as they are of increasing interest as potential therapeutic targets in heart failure.

METHODS SUMMARY

Ca^{2+}-ATPase (SERCA1a), isolated and affinity purified from rabbit hind leg muscle or adenovirus infected COS7 cells, was crystallized by vapour diffusion in the absence of Ca^{2+} but in the presence of 40 mM $MgSO_4$ or $MgCl_2$. Crystal structures were determined by molecular replacement. For identifying interacting residues between SERCA1a and sarcolipin, Cys mutants of SERCA1a and NFL-sarcolipin (Flag tag attached to the N terminus of sarcolipin)[7] were expressed in HEK293 cells; crosslinking was induced using copper phenanthroline or dibromobimane, and probed with anti-Flag monoclonal antibody, based on the procedure for phospholamban[24,25].

Full Methods and any associated references are available in the online version of the paper.

Received 3 August 2012; accepted 11 January 2013.
Published online 3 March 2013.

1. Toyoshima, C. Structural aspects of ion pumping by Ca^{2+}-ATPase of sarcoplasmic reticulum. *Arch. Biochem. Biophys.* **476,** 3–11 (2008).
2. Møller, J. V., Olesen, C., Winther, A. M. & Nissen, P. The sarcoplasmic Ca^{2+}-ATPase: design of a perfect chemi-osmotic pump. *Q. Rev. Biophys.* **43,** 501–566 (2010).
3. Tadini-Buoninsegni, F., Bartolommei, G., Moncelli, M. R., Guidelli, R. & Inesi, G. Pre-steady state electrogenic events of Ca^{2+}/H^{+} exchange and transport by the Ca^{2+}-ATPase. *J. Biol. Chem.* **281,** 37720–37727 (2006).
4. Champeil, P., Gingold, M. P., Guillain, F. & Inesi, G. Effect of magnesium on the calcium-dependent transient kinetics of sarcoplasmic reticulum ATPase, studied by stopped flow fluorescence and phosphorylation. *J. Biol. Chem.* **258,** 4453–4458 (1983).
5. Liu, Y. *et al.* High-yield heterologous expression of wild type and mutant Ca^{2+} ATPase: characterization of Ca^{2+} binding sites by charge transfer. *J. Mol. Biol.* **391,** 858–871 (2009).
6. Odermatt, A. *et al.* Characterization of the gene encoding human sarcolipin (SLN), a proteolipid associated with SERCA1: absence of structural mutations in five patients with Brody disease. *Genomics* **45,** 541–553 (1997).
7. Odermatt, A. *et al.* Sarcolipin regulates the activity of SERCA1, the fast-twitch skeletal muscle sarcoplasmic reticulum Ca^{2+}-ATPase. *J. Biol. Chem.* **273,** 12360–12369 (1998).
8. MacLennan, D. H. & Kranias, E. G. Phospholamban: a crucial regulator of cardiac contractility. *Nature Rev. Mol. Cell Biol.* **4,** 566–577 (2003).
9. Periasamy, M., Bhupathy, P. & Babu, G. J. Regulation of sarcoplasmic reticulum Ca^{2+} ATPase pump expression and its relevance to cardiac muscle physiology and pathology. *Cardiovasc. Res.* **77,** 265–273 (2008).
10. Kranias, E. G. & Hajjar, R. J. Modulation of cardiac contractility by the phospholamban/SERCA2a regulatome. *Circ. Res.* **110,** 1646–1660 (2012).
11. Bal, N. C. *et al.* Sarcolipin is a newly identified regulator of muscle-based thermogenesis in mammals. *Nature Med.* **18,** 1575–1579 (2012).
12. Obara, K. *et al.* Structural role of countertransport revealed in Ca^{2+} pump crystal structure in the absence of Ca^{2+}. *Proc. Natl Acad. Sci. USA* **102,** 14489–14496 (2005).
13. Jensen, A. M., Sorensen, T. L., Olesen, C., Møller, J. V. & Nissen, P. Modulatory and catalytic modes of ATP binding by the calcium pump. *EMBO J.* **25,** 2305–2314 (2006).
14. Toyoshima, C. & Nomura, H. Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* **418,** 605–611 (2002).
15. Toyoshima, C., Nomura, H. & Tsuda, T. Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues. *Nature* **432,** 361–368 (2004).
16. Lübben, M. *et al.* Sulfate acts as phosphate analog on the monomeric catalytic fragment of the CPx-ATPase CopB from *Sulfolobus solfataricus*. *J. Mol. Biol.* **369,** 368–385 (2007).
17. Toyoshima, C., Yonekura, S., Tsueda, J. & Iwasawa, S. Trinitrophenyl derivatives bind differently from parent adenine nucleotides to Ca^{2+}-ATPase in the absence of Ca^{2+}. *Proc. Natl Acad. Sci. USA* **108,** 1833–1838 (2011).
18. Holloway, C. E. & Melnik, M. Magnesium compounds: classification and analysis of crystallographic and structural data. *J. Organomet. Chem.* **465,** 1–63 (1994).
19. Cantilina, T., Sagara, Y., Inesi, G. & Jones, L. R. Comparative studies of cardiac and skeletal sarcoplasmic reticulum ATPases. Effect of a phospholamban antibody on enzyme activation by Ca^{2+}. *J. Biol. Chem.* **268,** 17018–17025 (1993).
20. Afara, M. R., Trieber, C. A., Ceholski, D. K. & Young, H. S. Peptide inhibitors use two related mechanisms to alter the apparent calcium affinity of the sarcoplasmic reticulum calcium pump. *Biochemistry* **47,** 9522–9530 (2008).
21. Bhupathy, P., Babu, G. J., Ito, M. & Periasamy, M. Threonine-5 at the N-terminus can modulate sarcolipin function in cardiac myocytes. *J. Mol. Cell. Cardiol.* **47,** 723–729 (2009).
22. Tupling, A. R. *et al.* Enhanced Ca^{2+} transport and muscle relaxation in skeletal muscle from sarcolipin-null mice. *Am. J. Physiol. Cell Physiol.* **301,** C841–C849 (2011).
23. Asahi, M. *et al.* Sarcolipin regulates sarco(endo)plasmic reticulum Ca^{2+}-ATPase (SERCA) by binding to transmembrane helices alone or in association with phospholamban. *Proc. Natl Acad. Sci. USA* **100,** 5040–5045 (2003).
24. Toyoshima, C. *et al.* Modeling of the inhibitory interaction of phospholamban with the Ca^{2+} ATPase. *Proc. Natl Acad. Sci. USA* **100,** 467–472 (2003).
25. Chen, Z., Akin, B. L., Stokes, D. L. & Jones, L. R. Cross-linking of C-terminal residues of phospholamban to the Ca^{2+} pump of cardiac sarcoplasmic reticulum to probe spatial and functional interactions within the transmembrane domain. *J. Biol. Chem.* **281,** 14163–14172 (2006).
26. Akin, B. L., Chen, Z. & Jones, L. R. Superinhibitory phospholamban mutants compete with Ca^{2+} for binding to SERCA2a by stabilizing a unique nucleotide-dependent conformational state. *J. Biol. Chem.* **285,** 28540–28552 (2010).
27. Bidwell, P., Blackwell, D. J., Hou, Z., Zima, A. V. & Robia, S. L. Phospholamban binds with differential affinity to calcium pump conformers. *J. Biol. Chem.* **286,** 35044–35050 (2011).
28. Buffy, J. J. *et al.* Defining the intramembrane binding mechanism of sarcolipin to calcium ATPase using solution NMR spectroscopy. *J. Mol. Biol.* **358,** 420–429 (2006).
29. Olesen, C., Sørensen, T. L., Nielsen, R. C., Møller, J. V. & Nissen, P. Dephosphorylation of the calcium pump coupled to counterion occlusion. *Science* **306,** 2251–2255 (2004).
30. Hayward, S. Structural principles governing domain motions in proteins. *Proteins* **36,** 425–435 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. Suzuki and S. Danko (Asahikawa Medical College) for providing us with much unpublished information. Thanks are also due to S. Hasegawa (JASRI) and H. Mimura for data collection at BL41XU of SPring-8, S. Yonekura for refinement of the E2 crystals and Y. Norimatsu for molecular dynamics simulations. We are grateful to D. B. McIntosh for help in improving the manuscript. This work is part of a long-term project (2009B0025) at SPring-8, and was supported by a Specially Promoted Project Grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to C.T.).

Author Contributions C.T. and G.I. planned and supervised the study; H.O. carried out all the DNA work; H.O., A.H., S.I. and J.T. performed protein preparation and other biochemical work; S.I. and C.T. crystallized the proteins; G.I. established the methods for large-scale production of recombinant proteins and characterized several mutants for clarifying the activation signal; C.T. and H.O. collected diffraction data and determined the structure; C.T., H.O. and J.T. prepared figures; and C.T. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures are deposited in the Protein Data Bank under accession codes 3W5D (E2+Pi), 3W5C (E2), 3W5A (E1·Mg^{2+}(native enzyme with bound sarcolipin)) and 3W5B (E1·Mg^{2+} (recombinant enzyme devoid of sarcolipin)). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.T. (ct@iam.u-tokyo.ac.jp).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

DNA and recombinant adenovirus construction. A complementary DNA library was created using rabbit skeletal muscle, using conventional methods. Then, the plasmid containing cDNA of *SERCA1a* (also known as *ATP2A1*) was selected and sequenced. Coding sequence for *SERCA1a* was amplified by PCR using oligonucleotide primers (sense: 5′-GCGGATCGCTAGCAGGGAGCGCA ATGGAG-3′, antisense: 5′-GGATCCGAATTCTTATCCCTCCAGGTAGTT CC-3′). Underlining represents restriction enzyme recognition sites. The PCR product was subcloned into NheI and EcoRI sites of the pIRES2-AcGFP1 vector (Clontech), which contains the internal ribosome entry site (IRES) and a sequence coding for a derivative of green fluorescent protein (AcGFP1). The resultant vector was named pIRES2-SERCA1a. Then *SERCA1a* including the IRES and AcGFP1 regions in pIRES2-SERCA1a was amplified by PCR using oligonucleotide primers (sense: 5′-GACCGCGATCGCCATGGAGGCCGCGCACTCCAA-3′, antisense: 5′-TTGTGTTTAAACTCACTTGTACAGCTCATCCATGCCGT-3′). The PCR product was subcloned into SgfI and PmeI sites of the pFN21K vector (Promega), fusing HaloTag and the tobacco etch virus (TEV) protease recognition site to the N terminus of SERCA1a. The resultant sequence of HaloTag-fused *SERCA1a* including IRES and AcGFP regions was amplified by PCR using oligonucleotide primers (sense: 5′-GCGGATCGCGGCCGCGCAAAGCCACCAT GGCAGA-3′, antisense: 5′-GGATCCGATATCTCACTTGTACAGCTCATC CATGCC-3′). The PCR product was subcloned into NotI and EcoRV sites of the pShuttle-CMV vector (Stratagene). AdEasy XL adenoviral vector system (Stratagene) was used to establish recombinant adenovirus.

Enzyme preparation. Native Ca^{2+}-ATPase (SERCA1a) was extracted from sarcoplasmic reticulum membrane with 2% octaethyleneglycol mono-*n*-dodecylether ($C_{12}E_8$) and purified by affinity chromatography as described[31]. For wild-type rabbit SERCA1a, COS7 cells (typically in 120 Petri dishes of 150 mm diameter) infected with recombinant adenovirus and cultured in DMEM supplemented with 5% FBS and 1% penicillin/streptomycin were collected, and the microsomal fraction prepared as described[5]. HaloTag-fused SERCA1a was solubilized using $C_{12}E_8$ and purified using the HaloTag protein purification system (Promega). SERCA1a was separated from HaloTag with TEV protease and purified further by dye affinity chromatography similarly to native enzyme.

Crystallization. Crystals were obtained by vapour diffusion at 10 °C and flash frozen in cold nitrogen gas. For E2 crystals, the protein solution contained 4 mg ml^{-1} affinity purified SERCA1a, 0.7% $C_{12}E_8$, 2.8 mg ml^{-1} phosphatidylcholine, 22% glycerol, 40 mM $MgSO_4$, 50 mM sodium butyrate, 3 mM EGTA, 0.1 mM dithiothreitol (DTT), 1 μg ml^{-1} butylhydroxytoluene (BHT) and 7 mM PIPES, pH 6.5. The reservoir consisted of 25% glycerol, 15% PEG6000, 40 mM $MgSO_4$, 150 mM sodium butyrate, 9 mM EGTA, 0.3 mM DTT, 3 μg ml^{-1} BHT and 7 mM PIPES, pH 6.8. The pH was adjusted with HCl after mixing the salts. E2 crystals were grown to 400 × 400 × 50 μm in about one month. For E1·Mg^{2+} crystals, the protein solution contained 4 mg ml^{-1} affinity purified SERCA1a, 0.3 mM TNPAMP, 3.2 mg ml^{-1} phosphatidylcholine, 0.7% $C_{12}E_8$, 15% glycerol, 4% PEG6000, 40 mM $MgSO_4$, 0.1 mM DTT, 100 mM acetic acid, 1 mM EGTA, 1 μg ml^{-1} BHT, and 13 mM triethanolamine, pH 7.5. The reservoir consisted of 23% glycerol, 10% PEG6000, 40 mM $MgSO_4$, 0.1 mM DTT, 150 mM acetic acid, 1 mM EGTA, 3 μg ml^{-1} BHT, 40 mM triethanolamine, pH 7.5. The pH was adjusted with *n*-methyl-D-glucamine after mixing to avoid introducing Na^+ and K^+ as far as possible. Crystals of native enzyme were grown to 400 × 100 × 20 μm, and those of recombinant enzyme to 300 × 50 × 20 μm in about one month.

Structure determination and analysis. All the diffraction data were collected at BL41XU of SPring-8 at 100 K using a Rayonix MX225HE CCD detector. The wavelength used was 0.9 Å. Diffraction data were processed with Denzo and Scalepack[32]. The unit cell parameters, diffraction data and refinement statistics of the crystals analysed are listed in Supplementary Table 1. Molecular replacement and refinement were carried out with CNS[33], and finally with REFMAC[34] if TLS refinement was required. Starting models used were that of E1·$2Ca^{2+}$ (PDB ID: 1SU4) for E1·Mg^{2+} and E2(TG+BHQ) (PDB ID: 2AGV) for E2 crystals. Initial coordinates of water molecules associated with Mg^{2+} in the atomic model of E1·Mg^{2+} were taken from that of E2·TNPADP (PDP ID: 3AR6) for the Mg^{2+} bound to TNPAMP, and from preliminary all-atom molecular dynamics simulation for the transmembrane Mg^{2+}. For molecular dynamics simulation, the NAMD software package[35] was used with CHARMM27 force field parameters[36] for proteins and ions, CHARMM36 for phospholipids[37] (dioleoylphosphatidylcholine), and TIP3 for water molecules. The quality of the model was examined with Procheck[38]. Ramachandran statistics (favoured, allowed, generously allowed, disallowed (%)) are: 84.5, 15.1, 0.4, 0 (E1·Mg^{2+} (native)); 81.5, 17.7, 0.8, 0 (E1·Mg^{2+} (recombinant)); 87.3, 12.3, 0.5, 0 (E2($MgCl_2$)); 88.9, 10.9, 0.2, 0 (E2($MgSO_4$)). Hydrogen bonds were identified with HBPLUS[39]. Domain movements were analysed with Dyndom[30]. Structural figures were prepared with TurboFRODO, Molscript[40] and PyMOL (The PyMOL Molecular Graphics System, Schrödinger).

Chemical crosslinking and detection of sarcolipin. The coding sequence of NFL–sarcolipin (rabbit sarcolipin with the Flag tag (MDYKDDDDK) attached to the N terminus)[7] was synthesized, amplified by PCR, subcloned into BamHI and EcoRI sites of pIRES2-AcGFP1 vector (Clontech) and transfected into HEK-293 cells by calcium phosphate methods. Cys-substitution, expression, and crosslinking experiments were carried out following refs 24 and 25. In short, microsome preparations from HEK-293 cells that co-expressed SERCA1a and NFL–sarcolipin were incubated in a buffer containing 0.15 M sucrose, 0.15 M KCl, 5 mM $MgCl_2$, 1.75 mM glutathione (only in experiments using copper phenanthroline), 2.5 mM EGTA and 25 mM triethanolamine-HCl, pH 7.5 (E1·$2Ca^{2+}$ and E1·Mg^{2+}) or 25 mM MES-NaOH, pH 6.0 (E2 and E2(TG)). For E1·$2Ca^{2+}$, 1 mM $CaCl_2$ was included instead of EGTA. For E2(TG), 20 μM TG was included. Disulphide bonds were formed by adding copper phenanthroline (0.3 mM $CuSO_4$ and 0.9 mM 1,10-phenanthroline) and incubating for 10 min at room temperature. For crosslinking with dibromobimane (bBBr), 0.16 mM bBBr in dimethylsulphoxide was added and reacted for 1 h. The reaction was stopped by addition of 25 mM *N*-ethylmaleimide. Proteins were separated on 4–12% Bis-Tris gradient gels (Invitrogen), transferred to PVDF (polyvinylidene difluoride) membrane and probed with anti-Flag monoclonal antibody (Sigma).

For detection of native sarcolipin, anti-sarcolipin polyclonal antibody was raised (Sigma) as described[41] (except that the N terminus of the sarcolipin C-terminal peptide was linked to haemocyanin). Interestingly, this antibody did not react with crosslinked products. Edman degradation was also performed for determination of the N-terminal sequence, which was MERSTRELXL (X corresponding to C, in this case), agreeing with that of sarcolipin. For quantification, sarcolipin was prepared from rabbit back muscle[42].

31. Coll, R. J. & Murphy, A. J. Purification of the CaATPase of sarcoplasmic reticulum by affinity chromatography. *J. Biol. Chem.* **259,** 14249–14254 (1984).
32. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
33. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
34. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D* **55,** 247–255 (1999).
35. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
36. MacKerell, A. D. Jr *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–3616 (1998).
37. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114,** 7830–7843 (2010).
38. Collaborative Computational Project, Number 4. The *CCP*4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
39. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238,** 777–793 (1994).
40. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
41. Babu, G. J., Bhupathy, P., Carnes, C. A., Billman, G. E. & Periasamy, M. Differential expression of sarcolipin protein during muscle development and cardiac pathophysiology. *J. Mol. Cell. Cardiol.* **43,** 215–222 (2007).
42. Ohnoki, S. & Martonosi, A. Purification and characterization of the proteolipid of rabbit sarcoplasmic reticulum. *Biochim. Biophys. Acta* **626,** 170–178 (1980).

©2013 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature11900

The sarcolipin–bound calcium pump stabilizes calcium sites exposed to the cytoplasm

Anne-Marie L. Winther[1,2,3]*, Maike Bublitz[2,3]*, Jesper L. Karlsen[2,3], Jesper V. Møller[2,4], John B. Hansen[1], Poul Nissen[2,3] & Morten J. Buch-Pedersen[1]

The contraction and relaxation of muscle cells is controlled by the successive rise and fall of cytosolic Ca^{2+}, initiated by the release of Ca^{2+} from the sarcoplasmic reticulum and terminated by re-sequestration of Ca^{2+} into the sarcoplasmic reticulum as the main mechanism of Ca^{2+} removal. Re-sequestration requires active transport and is catalysed by the sarcoplasmic reticulum Ca^{2+}-ATPase (SERCA), which has a key role in defining the contractile properties of skeletal and heart muscle tissue. The activity of SERCA is regulated by two small, homologous membrane proteins called phospholamban (PLB, also known as PLN) and sarcolipin (SLN)[1,2]. Detailed structural information explaining this regulatory mechanism has been lacking, and the structural features defining the pathway through which cytoplasmic Ca^{2+} enters the intramembranous binding sites of SERCA have remained unknown. Here we report the crystal structure of rabbit SERCA1a (also known as ATP2A1) in complex with SLN at 3.1 Å resolution. The regulatory SLN traps the Ca^{2+}-ATPase in a previously undescribed E1 state, with exposure of the Ca^{2+} sites through an open cytoplasmic pathway stabilized by Mg^{2+}. The structure suggests a mechanism for selective Ca^{2+} loading and activation of SERCA, and provides new insight into how SLN and PLB inhibition arises from stabilization of this E1 intermediate state without bound Ca^{2+}. These findings may prove useful in studying how autoinhibitory domains of other ion pumps modulate transport across biological membranes.

The regulation of SERCA by SLN and PLB is a key determinant of the contractile properties of both skeletal and heart muscle tissue. In humans, SLN is primarily expressed in skeletal muscle, whereas PLB expression predominates in the heart. SLN regulation is also proposed to be involved in non-shivering thermogenesis[3]. SLN and PLB are believed to bind to and regulate SERCA in a similar fashion. Binding of SLN or PLB to SERCA lowers the apparent Ca^{2+} affinity, thus modulating the threshold Ca^{2+} concentration at which SERCA gains transport activity[2–6]. Phosphorylation of SLN and PLB is a key regulatory mechanism of SERCA activity, as SERCA is only inhibited by the non-phosphorylated proteins at resting Ca^{2+} conditions[6,7].

The structures of SLN and PLB and their interactions with SERCA have previously been studied by NMR[8,9], but the structure of the complex of either of the regulators bound to SERCA has so far only been observed for PLB at low resolution by electron microscopy[10]. Numerous crystal structures have, on the other hand, provided a thorough description of SERCA at atomic detail[11] (for an account of prior structural studies, see Methods). The 110 kilodalton single polypeptide chain folds into a bundle of ten transmembrane segments (M1–M10) that form the ion-transport domain, and three cytoplasmic domains (A, P and N), which mediate ATP hydrolysis through phosphorylation and dephosphorylation. SERCA undergoes large conformational changes during the Ca^{2+} transport cycle: the E1 state binds Ca^{2+} and reacts with ATP to form the ADP-sensitive phosphoenzyme ($[Ca_2]$E1P), which occludes two Ca^{2+} ions at intramembranous sites I and II. Conversion to the ADP-insensitive E2P conformation exposes a luminal exit pathway, in which Ca^{2+} is exchanged for luminal protons. Dephosphorylation accompanies occlusion of two to three protons, and release of inorganic phosphate leads to the $[H_n]$E2 state. An open question remains, however, as to how the $[H_n]$E2 to $[Ca_2]$E1P transition occurs; the structure of a H^+/Ca^{2+}-exchanging E1 intermediate state is missing.

Rabbit SERCA1a (referred to hereafter as SERCA for simplicity) was solubilized from native sarcoplasmic reticulum membranes and crystallized in the absence of Ca^{2+}. The structure revealed a previously unknown conformation of SERCA with bound SLN (Fig. 1 and Supplementary Fig. 1). SLN forms a slightly curved membrane-spanning helix, and is associated with SERCA through a groove between M2, M6 and M9 (Fig. 1a). This is consistent with earlier docking models using a $[H_n]$E2 structure of SERCA[5]. However, the SLN-bound crystal structure is different from $[H_n]$E2 conformations (root mean squared deviation (r.m.s.d.) ~6.6 Å) and even though it is a Ca^{2+}-free state, it resembles more closely the $[Ca_2]$E1P-like structures (r.m.s.d. ~4.9 Å, encompassing the non-hydrolysable ATP analogue adenosine 5′-(β,γ-methylene)-triphosphate (AMPPCP)- and ADP–AlF_4^--bound forms, and the adenylyl imidodiphosphate (AMPPNP)-phosphorylated enzyme; Fig. 1b and Supplementary Figs 2 and 3).

Specifically, the positions of the cytoplasmic domains are E1-like, although they are not in a catalytically competent configuration as in the $[Ca_2]$E1P states (Fig. 1b and Supplementary Fig. 2). Electron density is observed for a partially ordered AMPPCP molecule, which overlaps with the position of the β,γ-phosphates and Mg^{2+} cofactor of bound AMPPCP in the $[Ca_2]$E1P state (Supplementary Fig. 3).

Compared to $[H_n]$E2 and $[Ca_2]$E1P, M1–M4 are shifted away from the central M6 in the SERCA–SLN complex (Supplementary Fig. 4). Notably, M1 is translocated into the membrane by almost two turns relative to M4, ~12 Å along the membrane normal (Fig. 2c, d). This results in an opening of a pathway between the cytoplasm and the Ca^{2+}-binding sites above the characteristic M1 kink at Leu 61, and a distorted membrane interface (Fig. 2a). Glu 309 has a key role in Ca^{2+} binding and release[11,12]. The closely located conserved pair of the hydrophobic Leu 61 (M1) and small Gly 257 (M3) has been described as a 'pivot point' for the kink, and the conserved Leu 65 as a 'gate-lock' residue[13], restricting Glu 309 mobility, which is required for ion occlusion. However, in the SERCA–SLN complex, Leu 61 faces Ala 303 (M4), Leu 65 packs against Val 300 (all conserved), and as a 'sliding door' the M1 kink translocates from 'above' to 'below' (luminal side down) the unwound, central part of M4 around the conserved Pro 308 and Glu 309. This imposes a deep distortion of the membrane interface, through which Glu 309 and other Ca^{2+}-binding residues are accessible from the cytoplasmic environment (Fig. 2a–d and Supplementary Videos 1 and 2). Accordingly, we consider the SLN

[1]Pcovery, Thorvaldsensvej 57, DK-1871 Frederiksberg, Denmark. [2]Centre for Membrane Pumps in Cells and Disease—PUMPkin, Danish National Research Foundation, DK-8000 Aarhus C, Denmark. [3]Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10C, DK-8000 Aarhus, Denmark. [4]Department of Biomedicine, Aarhus University, Ole Worms Allé 6, DK-8000 Aarhus, Denmark.
*These authors contributed equally to this work.

©2013 Macmillan Publishers Limited. All rights reserved

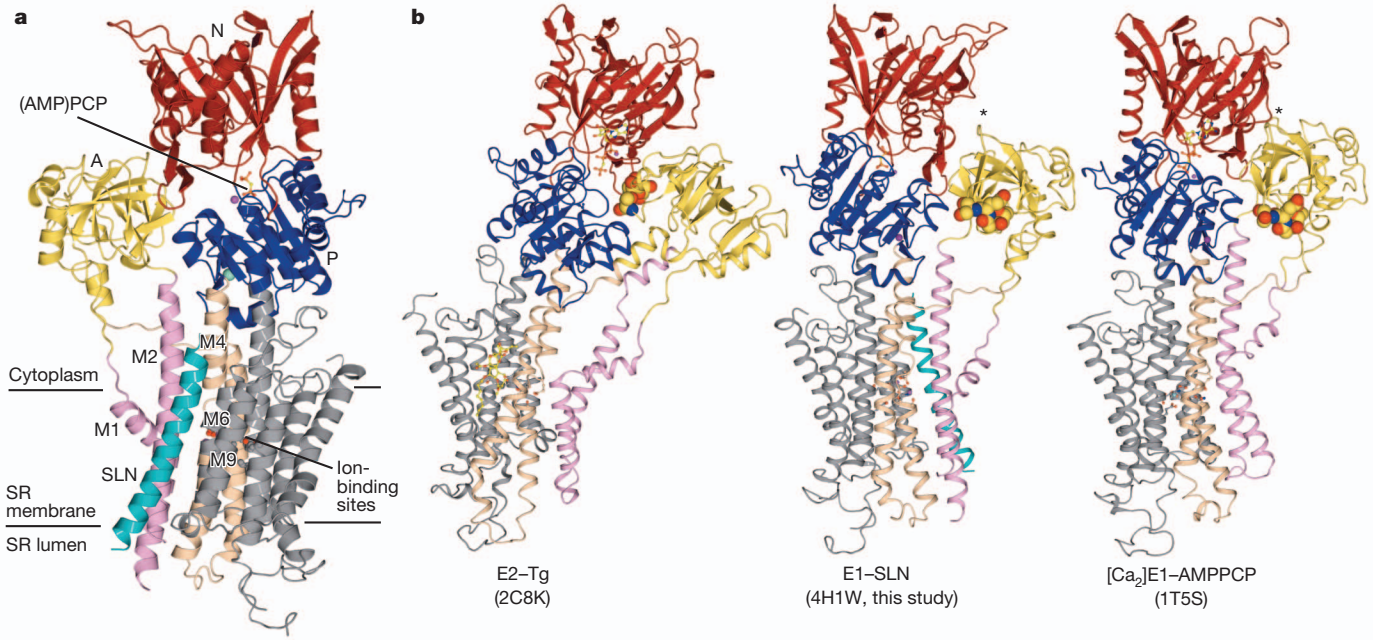


Figure 1 | Structure of the SERCA–SLN complex. **a**, Overall cartoon model of the SERCA–SLN complex showing the SLN-binding site between transmembrane helices M2, M6 and M9 of SERCA. Mg^{2+} is shown as magenta spheres and K^+ as a green sphere. The β- and γ-phosphates of AMPPCP and the Ca^{2+}-binding-site residues are shown as ball and stick. SR, sarcoplasmic reticulum. **b**, Comparison of the SLN-bound SERCA structure (middle) with representatives of the $[H_n]$E2 (left, stabilized by thapsigargin (Tg) and AMPPCP) and $[Ca_2]$E1P (right, stabilized by Ca^{2+} and AMPPCP) conformations. The hallmark Thr-Gly-Glu-Ser (TGES) motif in the A-domain (yellow) is highlighted as spheres. The arrangement of the cytoplasmic domains of the SERCA–SLN structure is E1-like, whereas transmembrane helices M1–M4 adopt a unique configuration. Asterisks mark a contact between the N-domain (red) and the A-domain (yellow), which is characteristic for $[Ca_2]$E1P states but absent in E1–SLN. Protein Data Bank (PDB) codes are given in parentheses below each structure.

complex to represent the genuine E1 intermediate of SERCA (hereafter referred to as E1–SLN) that allows cytoplasmic ion exchange in transition from the proton-occluded $[H_n]$E2 to the calcium-occluded $[Ca_2]$E1P state, reciprocating the E2P state that performs Ca^{2+}/H^+ exchange on the luminal side[11].

Kinetic evidence points to a sequential and cooperative mechanism of Ca^{2+} binding to SERCA, in which binding of the first Ca^{2+} ion leads to an increased affinity for the second[14]. In further support of this kinetic model, single mutations at site I prevent binding of Ca^{2+} to both sites I and II, whereas single mutations at site II only interfere with Ca^{2+} binding to that site[15,16].

With the M1 sliding door open in SERCA–SLN, the side chains of Asp 800 and Glu 309, in addition to the main-chain carbonyl groups of Val 304 and Ile 307 of site II, are exposed to the surface of the pathway formed by M1, M3 and M4 (Fig. 2b). Several other highly conserved residues (Gln 56, Asn 101 and Gln 109) line the pathway, indicating that an ion-selective environment for Ca^{2+} recruitment and binding is exposed at the membrane interface. Indeed, a comparison of the electrostatic surface potential of the $[H_n]$E2, E1–SLN and $[Ca_2]$E1P conformations shows that the E1–SLN structure displays a deep, funnel-shaped and negatively charged path that leads to the Ca^{2+} sites (Supplementary Fig. 5). This limits energetic barriers for loading of the ion-binding sites located almost halfway through the membrane bilayer, and phospholipid head groups may further assist Ca^{2+} entry.

Probing the molecular contact surface for Ca^{2+} ions (ionic radius of 1.14 Å), the calcium-entry pathway leads from a distorted site II further on to site I (Fig. 2b). Interestingly, Asp 800 (sites I + II) and Glu 309 (site II) are associated with two Mg^{2+} ions in our structure, one of which is also in close vicinity to the exposed main-chain carbonyls of Val 304, Ala 305 and Ile 307 (Fig. 2b, f, g, Supplementary Fig. 6 and Supplementary Video 1). Initial Ca^{2+} binding at site I is likely to engage Asp 800, thus distorting the Mg^{2+} sites and preparing site II for Ca^{2+} binding as a basis for cooperativity. We assume that Mg^{2+} ions bound at low-affinity sites (here saturated at the ~75 mM concentration in the crystal) stabilize an open structure of site II with exposed carbonyl groups of the unwound part of M4 and deprotonated side chains of Glu 309 and Asp 800. In further support of a deprotonated E1 state, Glu 771 is separated from the hydrogen bond to Asn 796 observed in the $[H_n]$E2 conformation (Fig. 2e, f).

Kinetic studies are in support of two Mg^{2+} ions being able to access the ion-binding sites from the cytoplasmic side and also show that high Mg^{2+} inhibits Ca^{2+} binding[17,18]. The structure further supports a proposed 'gating function' of Mg^{2+} (ref. 19), and conceptually provides a basis for Mg^{2+} modulation of Ca^{2+}-binding kinetics that may delay SERCA activation relative to the contractile filaments.

The SLN molecule in the complex structure was traced in its complete 31 amino-acid length, and the model was validated by the overall quality of fit to the electron density maps. Amino- and carboxy-terminal residues are, however, poorly defined (Fig. 3a and Supplementary Fig. 7), and a role of the SLN C-terminal Tyr-Gln-Tyr motif in binding to the M1–M2 luminal loop[5] is therefore not directly confirmed (Supplementary Fig. 7). Most of the observed SERCA–SLN contacts (4 Å distance or less) involve highly conserved amino-acid residues (Supplementary Table 2), and the binding site agrees with previous observations and models based on mutational studies for both SLN and PLB[4,5,20–22]. The SERCA M4 segment bends at the unwound region of the ion-binding cavity within the membrane and crosses diagonally over the N-terminal end of SLN at the cytoplasmic interface. The intermolecular contacts, however, show only few intimate interactions acting as anchor points, namely SLN Thr 5 hydrogen bonded to SERCA Trp 932, SLN Asn 11 to SERCA Thr 805 and Gly 801 (Fig. 3b), and a hydrophobic cluster of SLN Val 19 and Ile 22 with Leu 953 of M9 (Fig. 3c, Supplementary Table 3 and Supplementary Video 1).

SLN is homologous to the membrane-spanning C-terminal segment of PLB (Supplementary Fig. 8), and the SERCA–SLN interactions we

©2013 Macmillan Publishers Limited. All rights reserved

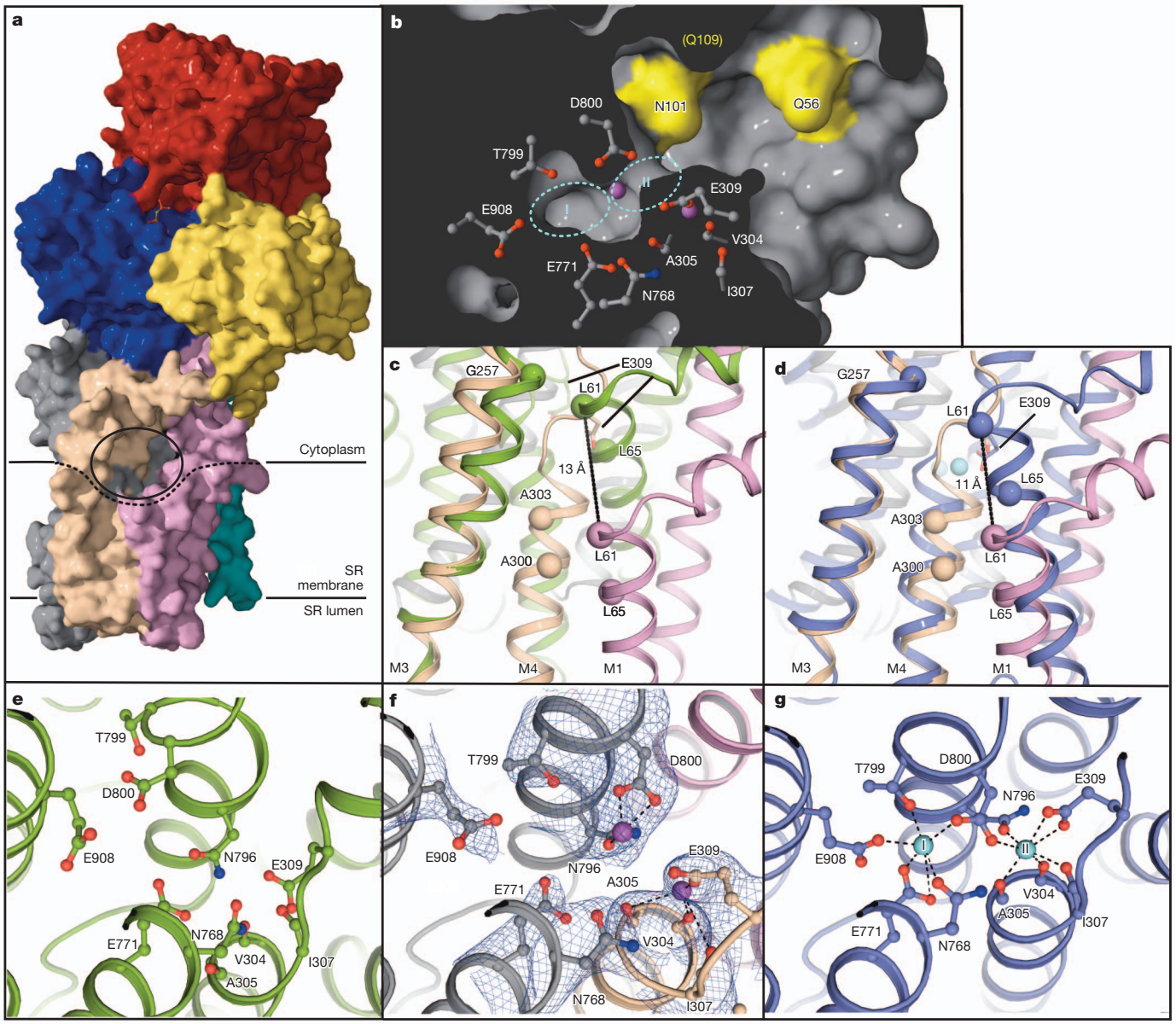


Figure 2 | Ca^{2+} access channel and binding sites. **a**, Surface representation of the SERCA–SLN complex: a wide funnel leads towards the Ca^{2+}-binding sites (circled). The 'M1 kink' is depressed into the membrane, indicating a distorted lipid interface (dotted line). Colouring as in Fig. 1a. **b**, Cross-section through the ion entry funnel. Yellow, conserved residues; magenta spheres, Mg^{2+} ions. Ca^{2+}-binding residues (ball and stick, Asn 796 omitted) of sites I and II (dotted circles) are projected on the cross-section. **c**, **d**, The M1 sliding door movement, with structures superimposed on M3. E2–Tg (PDB code, 2C8K) is in green (**c**) and [Ca_2]E1–AMPPCP (PDB code, 1T5S) is in blue (**d**). SERCA–SLN colouring as in **a**. **e**–**g**, The ion-binding sites. In the [H_n]E2 state both sites are distorted (**e**). In the E1–SLN state site II residues are involved in Mg^{2+} (magenta) binding and site I is accessible. Mesh represents refined $2F_o - F_c$ electron density map at 1.0σ (**f**). In the [Ca_2]E1P state two Ca^{2+} ions (cyan) are bound (**g**).

observe agree well with SERCA–PLB crosslinking studies[23] (Supplementary Fig. 9).

Regulatory activity by SLN and PLB is governed by their expression level and phosphorylation[2,4,6,7]. Both SLN and PLB have been proposed to exist in biological membranes in monomeric and multimeric free forms and as monomers and dimers in SERCA-bound forms[24]. Plotting the gain-of-function and loss-of-function mutations of PLB[4,21,25,26] onto the SLN structure, the mutations clearly segregate on opposite sides of the helix (Supplementary Fig. 10): loss-of-function mutations at the SERCA interface, and gain-of-function mutations (presumably impaired PLB pentamerization leading to more active monomer) on the outside.

This supports the notion that SERCA can interact, at least in part, with PLB pentamers[26], as the pentamerization-interface residues and SERCA-binding residues are on opposite sides, which may allow simultaneous formation of both PLB–PLB and SERCA–PLB interactions.

During β-adrenergic stimulation, PLB is phosphorylated by protein kinase A or Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) resulting in a rapid increase in Ca^{2+} uptake to the sarcoplasmic reticulum, which strongly affects relaxation and contraction[6]. SLN is also phosphorylated by CaMKII at Thr 5, and a phosphorylation mimic (Thr5Glu mutation) abolishes the inhibitory function of ectopically expressed SLN in adult rat ventricular myocytes, whereas a block in CaMKII activity in these cells eliminates a β-adrenergic response[7]. Thr 5 interacts with SERCA Trp 932, and phosphorylation at this site would cause a steric clash that destabilizes binding (Supplementary Fig. 11).

Both non-dissociative and dissociative models for SLN and PLB interaction with SERCA have been suggested[3,4,23]. Superimposing

©2013 Macmillan Publishers Limited. All rights reserved

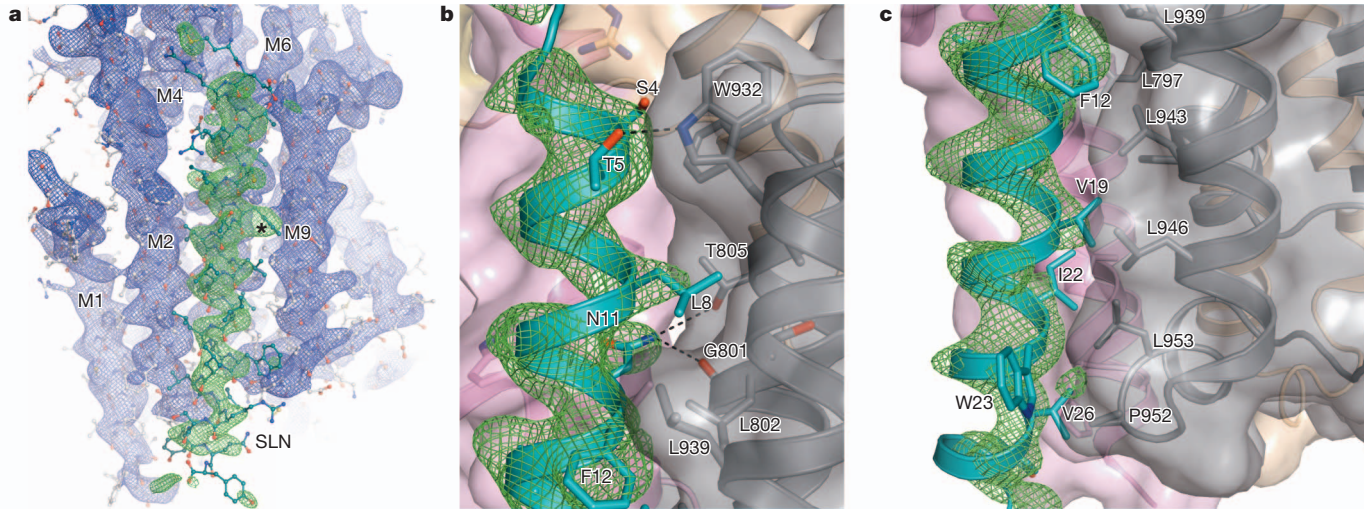


Figure 3 | SERCA–SLN interactions. **a**, Omit map showing the transmembrane α-helix formed by SLN. Green mesh represents unbiased $F_o - F_c$ difference density calculated at 50–3.1 Å resolution and contoured at 2.8σ, after simulated annealing refinement omitting the SLN model. The bulky-side-chain density of Phe 12 (asterisk) served as hallmark residue to assign the SLN sequence. Both the very N- and C-terminal residues are poorly defined. **b**, Close-up view of the polar interactions between the SLN N-terminal region and SERCA. Omit map as in **a**. **c**, Close-up view of hydrophobic contacts between the C-terminal region of SLN and SERCA. Omit map as in **a**.

our structure on the [H_n]E2 state of SERCA, we find a wider SLN-binding groove, that is, compatible with a SERCA [H_n]E2–SLN complex. It would offer a more extensive contact to M4 rather than to M2, and suggests a close contact to SERCA Leu 321, which when mutated reduces both SLN and PLB binding[4,5] (Fig. 4, left). In the occluded [Ca_2]E1P conformation, M1, M2, M3 and M4 have moved, narrowing the SLN-binding site. In particular, residues Leu 96, Trp 107 and Glu 108 constrict the groove for the N-terminal region of SLN with the anchor residues Thr 5 and Asn 11 (Fig. 4, right). We therefore think it is probable that SLN will be either dislodged from its binding site or at least undergo a positional rearrangement when SERCA shifts to the [Ca_2]E1P conformation (Supplementary Video 3). A recent study showing Ca^{2+}-resistant crosslinking of an SLN Glu7Cys mutant to SERCA as opposed to the analogous PLB Asn30Cys mutant complies with the latter, and thus suggests an ability of SLN to remain attached to the pump throughout the enzymatic cycle.

Many P-type ATPases are controlled by autoregulatory domains, such as the plasma membrane Ca^{2+}-ATPases[27,28] and heavy-metal pumps[29]. Plasma membrane Ca^{2+}-ATPases are activated by Ca^{2+}/calmodulin binding to their autoinhibitory domain, which as a straight helix has been mapped to a conserved groove[28] that appears as the cytoplasmic extension of the SLN-binding site (Supplementary Fig. 12).

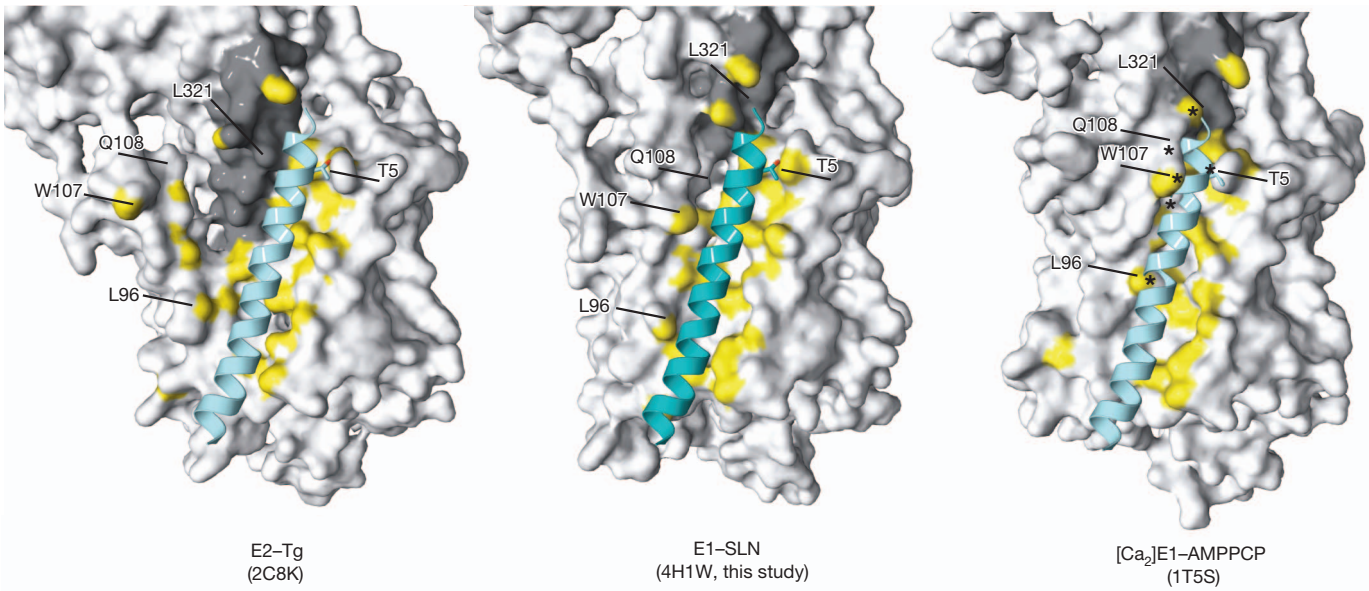


Figure 4 | Model of the regulation mechanism by SLN. Superposition of SLN in the E1–SLN structure (middle) and as modelled on the [H_n]E2 (left) and [Ca_2]E1P (right) states. Superposed SLN is shown in light cyan colour, with residues within 4 Å distance to SLN in the complex structure highlighted in yellow. The binding mode of SLN appears compatible with the [H_n]E2 state, with M4 (dark grey) being in more extensive contact to the SLN helix than in the other states. The mutational effect of Leu 321 on PLB/SLN binding can also be explained from the [H_n]E2 superposition. By contrast, tight SLN binding would be incompatible with the [Ca_2]E1P conformation, where the groove is constricted by SERCA residues Leu 96, Trp 107 and Gln 108, leading to clashes (asterisks) with SLN. PDB codes are given in parentheses below each structure.

©2013 Macmillan Publishers Limited. All rights reserved

The FXYD subunit of Na^+,K^+-ATPase is bound at the periphery of M9 in the $[K_2]E2.P_i$ state (see Methods), but may translocate to the M2–M6–M9 groove in an E1 state, in accordance with crosslinking data[30]. Taken together, these results hint at an auxiliary, regulation-sensitive surface present in P-type ion pumps that shifts exposure between E2 and E1 states, which could prove to be a favourable target site for therapeutic intervention.

METHODS SUMMARY

The SERCA1a–SLN complex was obtained from rabbit hind leg muscle. The protein was prepared in a Ca^{2+}-free buffer containing 1% dimethylsulphoxide and solubilized with the detergent octaethylene glycol monododecyl ether ($C_{12}E_8$). The crystals were obtained using PEG6000 as the precipitant and with approximately 75 mM $MgSO_4$ in the equilibrated crystallization drop. Cryoprotection was obtained by addition of a cryobuffer with 20% glycerol. Crystallographic data were collected at the beam lines 911-3 at MaxLab and X06SA at the Swiss Light Source. Initial phases were determined by molecular replacement using individual domains of SERCA as search models. Model building and refinement using data extending to 3.1 Å resolution produced a model with a crystallographic *R* factor of 19.1% and a free *R* factor of 24.6% (Supplementary Table 1).

Full Methods and any associated references are available in the online version of the paper.

Received 10 September 2012; accepted 11 January 2013.
Published online 3 March 2013.

1. James, P., Inui, M., Tada, M., Chiesi, M. & Carafoli, E. Nature and site of phospholamban regulation of the Ca^{2+} pump of sarcoplasmic reticulum. *Nature* **342,** 90–92 (1989).
2. Odermatt, A. *et al.* Sarcolipin regulates the activity of SERCA1, the fast-twitch skeletal muscle sarcoplasmic reticulum Ca^{2+}-ATPase. *J. Biol. Chem.* **273,** 12360–12369 (1998).
3. Bal, N. C. *et al.* Sarcolipin is a newly identified regulator of muscle-based thermogenesis in mammals. *Nature Med.* **18,** 1575–1579 (2012).
4. Asahi, M., Kimura, Y., Kurzydlowski, K., Tada, M. & MacLennan, D. H. Transmembrane helix M6 in sarco(endo)plasmic reticulum Ca^{2+}-ATPase forms a functional interaction site with phospholamban. Evidence for physical interactions at other sites. *J. Biol. Chem.* **274,** 32855–32862 (1999).
5. Asahi, M. *et al.* Sarcolipin regulates sarco(endo)plasmic reticulum Ca^{2+}-ATPase (SERCA) by binding to transmembrane helices alone or in association with phospholamban. *Proc. Natl Acad. Sci. USA* **100,** 5040–5045 (2003).
6. Tada, M. & Kadoma, M. Regulation of the Ca^{2+} pump ATPase by cAMP-dependent phosphorylation of phospholamban. *Bioessays* **10,** 157–163 (1989).
7. Bhupathy, P., Babu, G. J., Ito, M. & Periasamy, M. Threonine-5 at the N-terminus can modulate sarcolipin function in cardiac myocytes. *J. Mol. Cell. Cardiol.* **47,** 723–729 (2009).
8. Buffy, J. J. *et al.* Defining the intramembrane binding mechanism of sarcolipin to calcium ATPase using solution NMR spectroscopy. *J. Mol. Biol.* **358,** 420–429 (2006).
9. Lamberth, S. *et al.* NMR solution structure of phosphorlamban. *Helv. Chim. Acta* **83,** 2141–2152 (2000).
10. Young, H. S., Jones, L. R. & Stokes, D. L. Locating phospholamban in co-crystals with Ca^{2+}-ATPase by cryoelectron microscopy. *Biophys. J.* **81,** 884–894 (2001).
11. Møller, J. V., Olesen, C., Winther, A. M. & Nissen, P. The sarcoplasmic $Ca2^+$-ATPase: design of a perfect chemi-osmotic pump. *Q. Rev. Biophys.* **43,** 501–566 (2010).
12. Inesi, G., Ma, H., Lewis, D. & Xu, C. Ca^{2+} occlusion and gating function of Glu309 in the ADP-fluoroaluminate analog of the Ca^{2+}-ATPase phosphoenzyme intermediate. *J. Biol. Chem.* **279,** 31629–31637 (2004).
13. Einholm, A. P., Andersen, J. P. & Vilsen, B. Roles of transmembrane segment M1 of Na^+,K^+-ATPase and Ca^{2+}-ATPase, the gatekeeper and the pivot. *J. Bioenerg. Biomembr.* **39,** 357–366 (2007).
14. Dupont, Y. Low-temperature studies of the sarcoplasmic reticulum calcium pump. Mechanisms of calcium binding. *Biochim. Biophys. Acta* **688,** 75–87 (1982).
15. Zhang, Z. *et al.* Detailed characterization of the cooperative mechanism of Ca^{2+} binding and catalytic activation in the Ca^{2+} transport (SERCA) ATPase. *Biochemistry* **39,** 8758–8767 (2000).
16. Andersen, J. P. Dissection of the functional domains of the sarcoplasmic reticulum Ca^{2+}-ATPase by site-directed mutagenesis. *Biosci. Rep.* **15,** 243–261 (1995).
17. Smith, G. A., Vandenberg, J. I., Freestone, N. S. & Dixon, H. B. The effect of Mg^{2+} on cardiac muscle function: is CaATP the substrate for priming myofibril cross-bridge formation and Ca^{2+} reuptake by the sarcoplasmic reticulum? *Biochem. J.* **354,** 539–551 (2001).
18. Peinelt, C. & Apell, H. J. Kinetics of the Ca^{2+}, H^+, and Mg^{2+} interaction with the ion-binding sites of the SR Ca-ATPase. *Biophys. J.* **82,** 170–181 (2002).
19. Henderson, I. M., Starling, A. P., Wictome, M., East, J. M. & Lee, A. G. Binding of Ca^{2+} to the Ca^{2+}-Mg^{2+}-ATPase of sarcoplasmic reticulum: kinetic studies. *Biochem. J.* **297,** 625–636 (1994).
20. Toyoshima, C. *et al.* Modeling of the inhibitory interaction of phospholamban with the Ca^{2+} ATPase. *Proc. Natl Acad. Sci. USA* **100,** 467–472 (2003).
21. Toyofuku, T., Kurzydlowski, K., Tada, M. & MacLennan, D. H. Amino acids Glu2 to Ile18 in the cytoplasmic domain of phospholamban are essential for functional association with the Ca^{2+}-ATPase of sarcoplasmic reticulum. *J. Biol. Chem.* **269,** 3088–3094 (1994).
22. Kimura, Y., Asahi, M., Kurzydlowski, K., Tada, M. & MacLennan, D. H. Phospholamban domain Ib mutations influence functional interactions with the Ca^{2+}-ATPase isoform of cardiac sarcoplasmic reticulum. *J. Biol. Chem.* **273,** 14238–14241 (1998).
23. Chen, Z., Stokes, D. L., Rice, W. J. & Jones, L. R. Spatial and dynamic interactions between phospholamban and the canine cardiac Ca^{2+} pump revealed with use of heterobifunctional cross-linking agents. *J. Biol. Chem.* **278,** 48348–48356 (2003).
24. Mascioni, A., Karim, C., Barany, G., Thomas, D. D. & Veglia, G. Structure and orientation of sarcolipin in lipid environments. *Biochemistry* **41,** 475–482 (2002).
25. Kimura, Y., Kurzydlowski, K., Tada, M. & MacLennan, D. H. Phospholamban inhibitory function is activated by depolymerization. *J. Biol. Chem.* **272,** 15061–15064 (1997).
26. Glaves, J. P., Trieber, C. A., Ceholski, D. K., Stokes, D. L. & Young, H. S. Phosphorylation and mutation of phospholamban alter physical interactions with the sarcoplasmic reticulum calcium pump. *J. Mol. Biol.* **405,** 707–723 (2011).
27. Di Leva, F., Domi, T., Fedrizzi, L., Lim, D. & Carafoli, E. The plasma membrane Ca^{2+} ATPase of animal cells: structure, function and regulation. *Arch. Biochem. Biophys.* **476,** 65–74 (2008).
28. Tidow, H. *et al.* A bimodular mechanism of calcium control in eukaryotes. *Nature* **491,** 468–472 (2012).
29. Wu, C. C., Rice, W. J. & Stokes, D. L. Structure of a copper pump suggests a regulatory role for its metal-binding domain. *Structure* **16,** 976–985 (2008).
30. Füzesi, M. *et al.* Covalent cross-links between the gamma subunit (FXYD2) and α and β subunits of Na,K-ATPase: modeling the α-γ interaction. *J. Biol. Chem.* **280,** 18291–18301 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff at beamlines 911-3 at MaxLab and X06SA at the Swiss Light Source. We are grateful to N. Høgholm Jonassen, C. Olesen, B. Nielsen and A.-M. Nielsen for assistance with experimental procedures, and to H. Tidow and P. Gourdon for discussions on autoinhibition. We also thank P. Vangheluwe, H. Young, M. le Maire and E. Carafoli for discussions on PLB and SLN. A.-M.L.W. was supported by postdoctoral fellowships from the Danish Council for Independent Research (Technology and Production Sciences) and the Danish National Advanced Technology Foundation. P.N. is supported by an ERC advanced research grant (BIOMEMOS). Further support was obtained from the Fungalfight and Spotlight projects of the Danish Research Council for Strategic Research.

Author Contributions A.-M.L.W., J.B.H. and M.J.B.-P. initiated the project. J.V.M. provided sarcoplasmic reticulum membranes and purified SERCA1a, and A.-M.L.W. performed crystallization experiments. Data collection, structure determination and model refinement were performed by M.B. and A.-M.L.W. Videos were prepared by J.L.K. and A.-M.L.W. Model analysis and interpretation was done by M.B., P.N., A.-M.L.W., J.L.K., J.V.M. and M.J.B.-P. The paper was written by M.B., A.-M.L.W., M.J.B.-P. and P.N. with comments from all authors.

Author Information Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data Bank under accession number 4H1W. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.N. (pn@mb.au.dk) or M.J.B.-P. (mbp@pcovery.com).

©2013 Macmillan Publishers Limited. All rights reserved

METHODS

Protein preparation and crystallization. SERCA1a isoform was prepared from sarcoplasmic reticulum vesicles isolated from rabbit skeletal muscle. The membranes were further extracted and purified with a low concentration of deoxycholate according to established procedures[31]. The membrane preparation was washed in 100 mM MOPS-Tris, pH 6.8, and 80 mM KCl followed by a centrifugation at 40,000*g* for 35 min at 4 °C (TLS-55 rotor). The purified membranes were incubated for 80 min with 1% dimethylsulphoxide, 100 mM MOPS-KOH, pH 6.8, 20% glycerol, 80 mM KCl, 3 mM $MgCl_2$ and 3.9 mM EGTA. SERCA and SLN were solubilized with 35 mM octaethyleneglycol dodecylether ($C_{12}E_8$) for 10 min followed by centrifugation (TLA-55 rotor, 50,000*g*, 35 min, 4 °C). The protein concentration of the final supernatant was 12 mg ml^{-1} and 1 mM AMPPCP was added. Crystallization was obtained by vapour diffusion using hanging drops equilibrated at 23 °C from 1 + 1 µl protein and crystallization buffer (19.5% (w/v) PEG6000, 5% glycerol, 150 mM $MgSO_4$ and 6% 2-methyl-2,4-pentanediol (MPD). Under these conditions the equilibrated drops maintain approximately 2 µl in volume. Single crystals appeared after few days and grew within 1 week to a final size of 250 × 250 × 50 µm. Experiments initially aimed for co-crystallization with inhibitory compounds; however, these were not bound to the SERCA–SLN crystal structure presented here, and crystallization was later reproduced without inhibitors. Crystals were cryoprotected by addition of 1 µl cryobuffer (20% glycerol, 19.5% PEG6000, 150 mM $MgSO_4$ and 6% MPD) to a 2-µl drop. Crystals were mounted in LithoLoops and flash-cooled in liquid nitrogen.

Data collection and structure determination. Crystallographic data were collected at the beamlines 911-3 at MaxLab and X06SA at the Swiss Light Source (Supplementary Table 1). The data were collected at 100 K and a wavelength of 0.97935 Å. XDS was used for data processing[32] and phases were determined by molecular replacement using individual domains of SERCA (PDB codes, 2AGV and 3N5K) as search models with the program Phaser[33]. Sarcolipin was originally traced as a polyalanine model. Placing the 'distinct hallmark' side chain of SLN (Phe 12 in the central part of SLN) into the largest unbiased side-chain density yielded a reasonable fit of all other side chains in this helix register and satisfactory difference maps, whereas model refinement of alternative registers (shifted in both N- and C-terminal directions) increased the noise features of difference maps significantly. PHENIX was used for refinement[34], Coot for model building[35] and MolProbity for model validation[36]. The Ramachandran plot shows 93% of residues in favoured regions and 0.79% outliers (eight residues), with the latter emerging in poorly resolved loops upon reciprocal space refinement. Several structural reports on Ca^{2+}-ATPase, Cu^{+}-ATPase and Na^{+},K^{+}-ATPase were included in the model analysis[28,37–48] as well as atomic coordinates used for specific purposes (PDB codes, 1T5S, 2AGV, 2C8K and 3BA6).

31. Andersen, J. P., Lassen, K. & Møller, J. V. Changes in Ca^{2+} affinity related to conformational transitions in the phosphorylated state of soluble monomeric Ca^{2+}-ATPase from sarcoplasmic reticulum. *J. Biol. Chem.* **260,** 371–380 (1985).
32. Kabsch, W. *XDS. Acta Crystallogr. D* **66,** 125–132 (2010).
33. McCoy, A. J. *et al. Phaser* crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
34. Adams, P. D. *et al. PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
35. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
36. Arendall, W. B. *et al.* A test of enhancing model accuracy in high-throughput crystallography. *J. Struct. Funct. Genomics* **6,** 1–11 (2005).
37. Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405,** 647–655 (2000).
38. Toyoshima, C. & Nomura, H. Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* **418,** 605–611 (2002).
39. Sørensen, T. L., Moller, J. V. & Nissen, P. Phosphoryl transfer and calcium ion occlusion in the calcium pump. *Science* **304,** 1672–1675 (2004).
40. Toyoshima, C. & Mizutani, T. Crystal structure of the calcium pump with a bound ATP analogue. *Nature* **430,** 529–535 (2004).
41. Olesen, C., Sorensen, T. L., Nielsen, R. C., Moller, J. V. & Nissen, P. Dephosphorylation of the calcium pump coupled to counterion occlusion. *Science* **306,** 2251–2255 (2004).
42. Toyoshima, C., Nomura, H. & Tsuda, T. Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues. *Nature* **432,** 361–368 (2004).
43. Jensen, A. M., Sorensen, T. L., Olesen, C., Moller, J. V. & Nissen, P. Modulatory and catalytic modes of ATP binding by the calcium pump. *EMBO J.* **25,** 2305–2314 (2006).
44. Olesen, C. *et al.* The structural basis of calcium transport by the calcium pump. *Nature* **450,** 1036–1042 (2007).
45. Laursen, M. *et al.* Cyclopiazonic acid is complexed to a divalent metal ion when bound to the sarcoplasmic reticulum Ca^{2+}-ATPase. *J. Biol. Chem.* **284,** 13513–13518 (2009).
46. Toyoshima, C. How Ca^{2+}-ATPase pumps ions across the sarcoplasmic reticulum membrane. *Biochim. Biophys. Acta* **1793,** 941–946 (2009).
47. Gourdon, P. *et al.* Crystal structure of a copper-transporting PIB-type ATPase. *Nature* **475,** 59–64 (2011).
48. Morth, J. P. *et al.* Crystal structure of the sodium-potassium pump. *Nature* **450,** 1043–1049 (2007).

©2013 **Macmillan Publishers Limited. All rights reserved**

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature10977

Corrigendum: Modernization: One step at a time

Zhiguo Xu

Nature **480,** S90–S92 (2011)

A quote attributed to Yi Rao in this Outlook article incorrectly implied that he criticized the use of traditional Chinese medicine (TCM) techniques. Instead, Rao's intention was to question attitudes that apply different standards to TCM and modern medicine. The text 'This process is not helped by the insistence of many TCM researchers on using traditional methodologies — such as water decoction, immersion, reflux extraction and distillation — to process herbs. "It is totally ridiculous," says Rao' should instead read 'This process is not helped by the insistence of many TCM researchers on using different measures of efficacy for TCM compared with modern medicines. "You can use different drugs or approaches, but all drugs and treatment strategies have to pass the same standards to show that they are effective", says Rao'. This text has been replaced in the PDF and HTML versions online.

CORRIGENDUM

doi:10.1038/nature12022

Corrigendum: Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants

Wenqing Fu, Timothy D. O'Connor, Goo Jun, Hyun Min Kang, Goncalo Abecasis, Suzanne M. Leal, Stacey Gabriel, Mark J. Rieder, David Altshuler, Jay Shendure, Deborah A. Nickerson, Michael J. Bamshad, NHLBI Exome Sequencing Project & Joshua M. Akey

Nature **493,** 216–220 (2013); doi:10.1038/nature11690

In this Letter, Mark J. Rieder (Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA) was inadvertently omitted from the author list. He oversaw data generation and quality control and is a member of the Seattle Grand Opportunity group. The error has been corrected in the HTML and PDF versions of the paper online.

©2013 Macmillan Publishers Limited. All rights reserved

CAREERS

AUSTRALIA Plan to attract international students and researchers unveiled **p.273**

UNITED STATES Survey shows pay rise and perks for many university executives **p.273**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**



IKON IMAGES/CORBIS

NATURAL DISASTERS

A calculated risk

Scientists and engineers with an analytical bent are sought-after in natural-hazard risk assessment.

BY BRYN NELSON

Australian meteorologists introduced two new colours to their weather maps this January, during the country's hottest summer on record. Two years earlier, eastern Australia saw its worst flooding in decades following its wettest recorded December. Last autumn, Superstorm Sandy pummelled the US east coast, after devastating drought and wildfires scorched a huge swathe of the central states. Within the past three years, there have also been a catastrophic earthquake and tsunami in Japan, epic flooding in Thailand and East Africa's worst drought in 60 years. In a 2012 report, the Intergovernmental Panel on Climate Change concluded that climate change has already increased the frequency of some extreme weather events globally, and that more are on the way.

As these and other global threats weigh heavily on governments, corporations and charities, demand is spiking for scientists and engineers in the multidisciplinary field of natural-hazards risk analysis. Experts in the field, which focuses on how to predict, prevent and limit damage from natural disasters, are finding jobs in insurance, agriculture, finance, infrastructure, construction, humanitarian aid and public policy.

"There's a broad spectrum of the public and private sector that needs planning, that needs expertise, that needs a way of communicating the risks to effect policy," says Arthur Lerner-Lam, deputy-director of Columbia University's Lamont-Doherty Earth Observatory in Palisades, New York. "Good students, we're finding, are having no trouble finding positions in those areas." Particularly fruitful areas of study include Earth sciences, physical sciences and sustainable development, he says. "Some of our best PhD students are going to the insurance and finance sector, principally for modelling purposes."

Opportunities in risk analysis fall mainly into three overlapping categories: the natural and physical sciences; engineering; and social sciences. Engineers assess infrastructure and sometimes inform policy decisions; social scientists often study how best to communicate risk. Natural and physical scientists from geologists and meteorologists to mathematicians and physicists study the origins, movement and potential impacts of natural hazards. They often produce data-heavy models, maps, reports and briefs. The resulting visual presentations of risk may layer details such as a city's utility networks and roads onto a map of its flood-prone areas and a plot of the economic productivity of different zones as a function of their geography. These analyses help clients to estimate potential losses and to refine their prevention and mitigation measures.

BLURRING THE BOUNDARIES

A career in natural-hazards management and mitigation often means collaborating on multidisciplinary teams that may include meteorologists, economists and civil engineers. Employers say that all roles require a solid foundation in statistics and good analytical and quantitative skills. Some firms build websites to help their clients to visualize risk, so software-development experience and familiarity with tools such as geographic information systems can come in handy. And candidates must be able to dispense with jargon and explain what a risk might mean to a lay audience. Employers are increasingly looking for analysts from all disciplines with at least basic skills in risk communication, a growing area of study among sociologists and behavioural psychologists (see 'Reaching out — carefully').

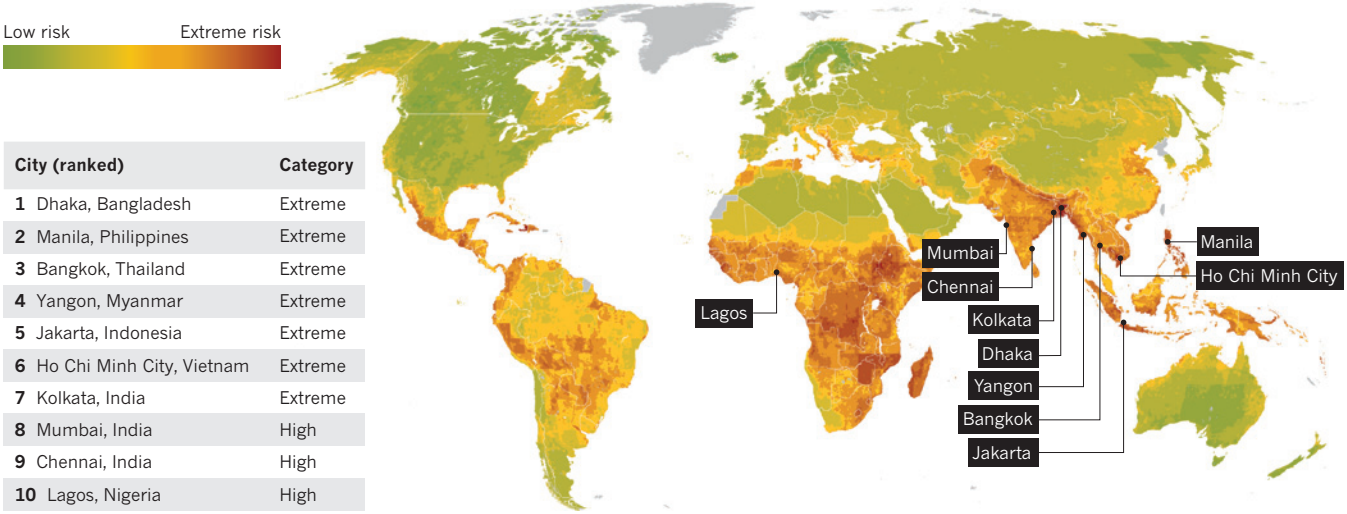
Jobs are particularly abundant at reinsurance companies, which insure other insurers to help them to limit their losses in the event of a disaster. Reto Schneider, head of emerging risk management at Swiss Re, a global reinsurance company based in Zurich, cites a long list of positions suitable for experts on disasters such as floods, droughts, windstorms and earthquakes, or for geographers and mathematicians who can do in-house modelling and mapping. For example, Swiss Re has hired several scientists from the Swiss Federal Institute for Snow and Avalanche Research in Davos to assist in risk analysis for landslides and avalanches. Consultancies that specialize in mapping or ▶

© 2013 Macmillan Publishers Limited. All rights reserved

DANGER DEPICTED

Risk analysts develop visual representations of risk levels, such as this 2013 map from Maplecroft in Bath, UK, which shows the cities most vulnerable to the effects of climate-related natural disasters and rising sea levels.

SOURCE: MAPLECROFT



City (ranked)	Category
1 Dhaka, Bangladesh	Extreme
2 Manila, Philippines	Extreme
3 Bangkok, Thailand	Extreme
4 Yangon, Myanmar	Extreme
5 Jakarta, Indonesia	Extreme
6 Ho Chi Minh City, Vietnam	Extreme
7 Kolkata, India	Extreme
8 Mumbai, India	High
9 Chennai, India	High
10 Lagos, Nigeria	High

modelling a client's vulnerability to a range of catastrophes are also big employers.

Schneider earned a PhD in immunology and did two postdoctoral fellowships before his search for a stable, family-friendly career led him to Swiss Re 18 years ago. The company was searching for what it now calls high potentials: clever and motivated academics who can be trained in an internal graduate programme. Schneider's current work in analysing the financial impact of disasters is far removed from immunology, but he sees parallels between the immune system and the complex systems that he now studies, and says that his analytical skills are an asset. "Scientists might overlook it or ignore it because they have never heard about it or they can't imagine that the insurance industry is offering something to them," he says. "But after close to 20 years working in this industry, I must say that it's a great target."

Helen Hodge is head of maps and indices at Maplecroft, a consultancy in Bath, UK, that produces maps, indices, scorecards and other visual representations of risk in ten major categories (see 'Danger depicted'). She estimates that half of her firm's recent recruits have scientific backgrounds. She has a master's degree in geography and worked as an environmental scientist before she started modelling natural catastrophes in London's insurance industry and eventually joined Maplecroft.

At Maplecroft, scientists are particularly well represented in risk analysis for natural hazards, climate change and environment, areas in which the 90-employee firm is scaling up its work. The expansion, says Hodge, is the result of demand from clients such as banks, insurance companies, major corporations and national and international humanitarian groups. "We're growing at a good rate," she says.

Commercial clients are increasingly wary of major disruptions in their supply chains, especially as many boost their reliance on supplies and parts from Southeast Asia, a region prone to natural disasters. Thailand's catastrophic monsoon flooding in 2011, in particular, delivered a sobering wake-up call about the danger of concentrating resources in a single location. Car and computer-hardware manufacturers were inundated: Honda and Toyota paused production and hard-drive prices soared by nearly 30%. Analysis of a company's business practice to identify high-risk locations or methods may help it to avoid the most vulnerable sites or to safeguard against lengthy delays in re-establishing its supply chain.

Experts in sustainability and resiliency are also finding work. Resiliency refers to designing buildings, infrastructure, contingency plans and communication systems that allow companies or communities to withstand a disaster or bounce back quickly. Sustainability experts help clients to make durable and cost-effective decisions about infrastructure, development and other long-term investments in the face of changing economic, climatic and political conditions (see *Nature* **494,** 507–509; 2013).

Gauging the integrity of infrastructure threatened by weather and seismic events is providing another big opportunity for scientists and engineers. For example, many of the 3,000 levees and 700 dams overseen by the US Army Corps of Engineers are reaching the end

COMMUNICATION SKILLS

Reaching out — carefully

Demand is rising for scientists who can not only assess risk but also clearly communicate it. "A lot of risk communication is trying to understand how people make decisions about risks, how they should make decisions about risks, how you help them make better decisions about risks," says George Gray, president of the 2,000-member Society for Risk Analysis in McLean, Virginia, an international organization of academic, government and private-sector professionals.

The engineering and public-policy department at Carnegie Mellon University in Pittsburgh, Pennsylvania, includes experimental psychologists who focus on developing, testing and gauging the response to risk-related messages aimed at policy-makers and the public. "You can sit in your office until the cows come home, trying to dream up appropriate messages, and still be surprised the minute you actually take them out in the field and test them," says Granger Morgan, head of the department. Even the best risk analysis in the world may be useless, he says, if it is not communicated in a way that improves how people perceive and respond to risk, and that supports good decision-making.

There is particular demand for scientists who can distil sophisticated information into understandable advice that influences behaviour — the likelihood of a levee withstanding a storm surge and how that affects a city's evacuation routes, for example — says Eric Halpin, special assistant for dam and levee safety at the US Army Corps of Engineers in Washington DC. "It's a balance of being a good-enough communicator, saying, 'This is why you should care about the risk,' without throwing a whole community into panic." **B.N.**

© 2013 Macmillan Publishers Limited. All rights reserved

of their typical 50-year life expectancy, and the costly process of shoring up or replacing them could take decades. "It's a long-term problem, so we need to make sure that we have the right people around 20 years from now to deal with this," says Eric Halpin, special assistant for dam and levee safety for the corps in Washington DC.

Halpin needs scientists and engineers who can assess more than just the logistics of building and maintaining dams and levees. They also need to understand how decisions about infrastructure may affect the long-term well-being of a local economy and population. "What we're finding out right now is that engineers tend to look at every problem as an opportunity to build something new, a structural solution, and many times the solutions are not structural," he says. "Maybe it's managing your consequences better through building codes, or smart economic development or evacuation plans rather than a bigger levee."

RISKY POLICIES

In the aftermath of devastating storms such as Sandy and 2005's Hurricane Katrina, and amid growing concern over the coastal impacts of global warming, government officials around the world are debating when to rebuild barriers and when to pursue other risk-reduction strategies (see D. Moynihan *Nature* **495,** 7; 2013). The increasingly urgent discussions are opening doors for scientists and engineers versed in public policy.

Frauke Hoss, who is doing a PhD in engineering and public policy at Carnegie Mellon University in Pittsburgh, Pennsylvania, is German but trained in political science and hydraulic engineering in the Netherlands. She hopes to pursue an advisory role with the European Union or the Dutch or German government, combining her expertise in public policy and flood-control engineering. Hoss says that infrastructure decisions are affected by geography in much of Europe. "We live so densely that you will be building in an area where there's flooding potential," she says. "And as your economy grows, you have more value that you have to keep drier."

In the Netherlands, where 35% of the population lives on flood-prone land, officials are supplementing flood barriers with new and larger spillways and reservoirs as more natural ways to control inundations. Although unsure of her own prospects in helping to formulate policies, Hoss thinks that she is on the right track with her dual focus. "There will always be demand, because there are very few engineers who are interested in policy," she says.

That demand is relevant across Europe. The UK government's Foresight programme, for example, relied heavily on natural-sciences and public-policy expertise in crafting its 139-page report *Reducing Risks of Future Disasters*, released last November. The programme commissioned 18 independently peer-reviewed papers and workshop reports and enlisted a small army of academic, governmental and institutional experts to write and review the report, which calls for decision-makers worldwide to increase their reliance on science, technology and risk-assessment methods to help to reduce the impact of disasters.

SLOW GROWTH IN ACADEMIA

Because risk analysis often spans disciplines and tends to be highly applied, academic positions are the hardest to find. However, a modest corps of academic sites focuses at least in part on natural hazards; most, but not all, are in the United States and Europe. Some of the best known established sites include the Aon Benfield UCL Hazard Centre at University College London; the Disaster Prevention Research Institute at Kyoto University in Japan; the Center for Disaster Management and Risk Reduction Technology in Potsdam, Germany; and the Engineering and Public Policy department at Carnegie Mellon. That list is now expanding, with centres such as the Cabot Institute at the University of Bristol, UK, opened in 2010, and the Center for Risk Governance at Tsinghua University in Beijing, launched this year.

Despite the growing options for academics and students, employers tend to agree that few programmes prepare scientists for the multidisciplinary, holistic and hands-on responsibilities of applied risk analysis. Applicants may have to actively seek out practical experience on their own, as Hodge did in the insurance industry or as others have done in organizations that regularly assess disasters, such as one of the many national Red Cross and Red Crescent societies.

Depending on the employer, recruits may receive extensive in-house training that directs their expertise to applied ends — Swiss Re, for example, sometimes trains employees in financial modelling, and the US Army Corps of Engineers teaches them how to incorporate skills beyond their normal disciplines. "We hire some straight-up scientists in hydrology and geology and geomorphology; we hire some straight-up social scientists, particularly in the areas of life loss and vulnerability; and we hire some straight-up engineers," says Halpin. "And not one of them comes to the table with all of the skills. We build it in after they get on board."

The aim is to produce what Halpin calls a natural-hazards "super generalist" who is well versed in the application of multiple disciplines to real-world problem solving. "If you have the right foundation in one area and you're willing to learn about other areas, and you're willing to create some of the education in the workplace rather than in school, then that's the key," he says. ■

Bryn Nelson *is a freelance writer based in Seattle, Washington.*

AUSTRALIA

Foreign students wanted

Hoping to boost Australia's economy by bringing in more international students and researchers, the nation's government is adopting recommendations from its International Education Advisory Council. On 27 February, the council published *Australia — Educating Globally*, a report that outlines a 5-year plan for reaching Australia's international-enrolment goals. The number of international students at universities and other institutions in Australia could grow by some 30% by 2020, to 520,000 students, as a result of the council's plan, the report predicts; it estimates that this would add about Aus$19 billion (US$19.5 billion) to the country's economy and provide extra support for research positions. The recommendations include streamlining visa requirements and regulations for international academics and students; encouraging international research collaborations and partnerships; and offering incentives for top academics and doctoral students to come to Australia. Michael Chaney, chairman of the advisory council, says that the government is already implementing specific strategies to increase numbers of international scientists, such as establishing high-paying contracts for research posts.

UNITED STATES

Pay rise for presidents

US university presidents and academic officers, including deans, provosts and chancellors, saw their pay increase by a median of 2.3% in 2012, more than one-third above the inflation rate of 1.7%, according to a survey. The College and University Professional Association for Human Resources in Knoxville, Tennessee, polled 1,251 institutions for its 2012–13 *Administrators in Higher Education Salary Survey*, published on 25 February. The median pay for the president of an institution that grants doctorates rose by 2% for the year, to US$402,000. Deans of biological and life sciences at doctoral institutions also saw a pay rise: their median compensation climbed by 3% to $225,000 in 2012. Executive benefits included not only pay but also housing for 47.3% of university presidents, a car or car allowance for 51.6% and membership of a club such as a gym or airline for 31.6%, the survey says. More than one-quarter of presidents received deferred compensation such as pensions or stock options.

© 2013 Macmillan Publishers Limited. All rights reserved

THE LUCKY ONES

After the plague.

BY ANDREW DAVID THALER



JACEY

My legs creak as I climb the stairs to our meeting room. I lean against the wall to steady myself. I could have taken the lift, saved myself the pain, but I need to heal. I am lucky. I can still walk after the disease ravaged my body, but my legs are weak, my arms emaciated, my face scarred and hollow. I keep my body moving, to help it heal. That's why we are all here: to heal.

I don't know why I still come, though. I don't get much from these meetings. They used to be comforting, but now they're just tedious. The survivors, overcome with grief or anger or disgust, are more likely to descend into fits of rage than to open up to any of us about their experiences. They are fighters, they had to be. You didn't survive, uninfected, by being soft. They internalized everything. Many were so consumed with guilt that they couldn't continue. Survivor suicides are an almost daily occurrence now.

I don't remember any of it. *Really*. One moment my wife and I are lying in bed, coughing violently with a fever that hasn't broken in days; next I wake up, strapped to a hospital bed, my limbs atrophied, my body ruined. It took days for any of it to make sense. Months before I could come to terms with it. My home was destroyed, burned in one of the many fires that swept the city. I don't know where my wife is or what happened to her.

I sit down, silently. It's a mixed group, the survivors and the cured, to "promote healing". These meetings are supposed to be open and anonymous, but the survivors know who the cured are. They can tell by our broken, rotted bodies. Most of the survivors refuse to share their names. I used to bring a picture of my wife to the meetings, on the off chance that a survivor would recognize her, but they never want to look. It's hard to think of us as people again.

I leave the meeting early and alone. Those of us who were cured enjoy being alone with our thoughts. We have been so long without them.

I stop at the corner store to buy some food for Mrs Richmond. Mrs Richmond was a nurse when the outbreak started. She lost her husband and grandson to the plague. She is a tough, no-nonsense woman, and she never watched scary movies. While her neighbours were gunning down their friends and family, she quietly hid her loved ones in the basement, feeding them when she could. Over the years she saved dozens of infected people, including me, from the rampaging and increasingly desperate mob.

It couldn't have been easy. Once, she tells me, a former neighbour, then head of the local militia, caught wind of her operation. The confrontation ended after a two-day stand-off between Mrs Richmond and four or five militia men. She still won't tell us how it was resolved. She is tough.

I owe her my life. Taking care of her shopping is the least I can do.

Mr Richmond greets me as I turn onto his block. He was once a doctor. He and Mrs Richmond met at the hospital where the outbreak began. He retired decades ago, but still visited her for lunch every day. He was among the first to be infected and his body wasted away during those years, leaving him wheelchair-bound and barely able to speak. But his mind is sharp, and for us, that is all that matters.

I live with the Richmonds now, on a small folding cot in their basement. Even with no memories of it, that basement feels safe.

The TV in the living room is tuned to the news. Another mass suicide of survivors, unable to cope with the consequences of the cure. The bridge, this time.

It's hard to blame them. We, the infected, were insatiable. In our mindless stupor we tore the life from numerous survivors. The illusion was perfect. We had no consciousness. Our bodies rotted where they stood. For all the survivors knew, we were dead, walking corpses slowly decomposing. It must have been terrifying. Those who weren't infected labelled us 'zombies', and, with 50 years of popular culture backing them up, launched an eradication campaign against us. It played out like every movie of the genre: hordes of the 'undead' against a ragtag group of survivors. Overwhelming odds.

And then someone found a cure. We weren't undead, we were sick.

The survivors call us the lucky ones. We were cured and it was over. We have no memory of the horrible world the disease created. The great mass of humanity became mindless zombies, stalking the living through a dystopian hellscape.

The rest became monsters. ■

Andrew David Thaler *is a postdoctoral researcher studying the ecology and conservation of deep-sea hydrothermal vents. He writes for the popular marine-science blog Southern Fried Science (southernfriedscience.com).*

➲ NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

© 2013 Macmillan Publishers Limited. All rights reserved